Sécurisation des données

Traitement des données sensibles

Auteur : Fabrice Sabatier
Nom de l'organisation : UL/Loria/CNRS
Date de création : 28-02-2022
Date de modification : 28-02-2022
Mots-cléfs : Segmentation, chiffrement, archivage, contrôle d'accès

Table des matières

1 Liste des abbréviations

DMZ	Demilitarized Zone.
FEDER	Fonds européen pour le développement régional.
LHS	Laboratoire de Haute Sécurité.
NAS	Network Attached Storage.
SFTP	Secure File Transfer Protocol.

2 Le Laboratoire de Haute Sécurité du LORIA

Le Laboratoire de Haute Sécurité (LHS) a été construit en 2008-2009 par le centre INRIA Nancy-Grand Est et le LORIA avec l'aide du Fonds européen pour le développement régional (FEDER), de la région Lorraine, de la métropole du Grand Nancy et du ministère de l'enseignement supérieur et de la recherche.

FIGURE 1 – Le LHS

Placé dans un environnement clos avec un réseau Internet isolé et avec accès protégé par reconnaissance biométrique, le laboratoire offre un cadre technologique et réglementaire fiable pour la conduite de tests et d'opérations sensibles. Il a été conçu pour garantir la sécurité des données, des phénomènes et des équipements analysés. Les axes de travail sont la virologie (reconnaître les virus de demain), la supervision des réseaux (analyse et sécuriser les échanges) et la détection. Il offre un cadre sécurisé pour héberger la plateforme LOLA.

3 La plateforme Lola

La plateforme LOLA a été réalisée dès sa conception en prenant en compte le caractère sensible des données qui sont traitées par les différents algorithmes et équipes de recherche associées au projet.

3.1 Segmentation de l'architecture

Afin de garantir un accès restreint aux données, la plateforme est divisée en plusieurs secteurs comme illustré sur la figure 2 :

— la zone d'échange avec l'extérieur DMZ (sous réseau isolé, tampon entre internet et le réseau sécurisé) ;
— la zone de confiance contenant les données sensibles au sein de bases de données, ainsi que le « backend » de l'application de gestion de LOLA ;

— La zone de production ou sont exécutés les algorithmes de traitement des données ;
— La zone de sauvegarde contenant exclusivement des données chiffrées.

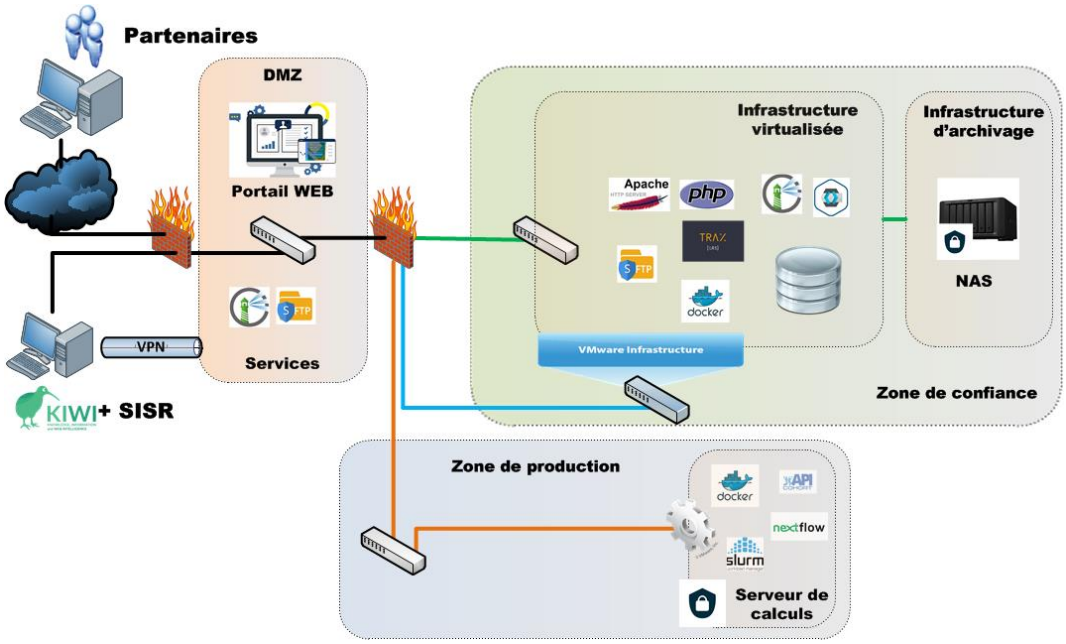

FIGURE 2 – Segmentation des zones au LHS

3.2 La zone d'échange dite DMZ

Dans cette zone comprise entre deux barrières de pare-feu de technologies différentes (Cisco et Juniper), se situe l'accès à l'interface web de l'application. Un service de reverse-proxy réalise la liaison filtrante entre l'utilisateur extérieur à la plateforme et le serveur web de l'application. On trouve également le service de dépôt des données offert aux partenaires et ainsi qu'un dépôt d'images Docker constituant un catalogue de scenarios. Chaque service s'exécute dans des machines virtuelles, permettant la création de sous-réseaux cloisonnés et la possibilité de redondance.

3.3 La zone de confiance

La zone de confiance se situe derrière les pares-feu. Les données (logs normalisés xAPI/TRAX et informations des utilisateurs) sont stockées dans des bases de données. Cette zone héberge les services permettant de faire fonctionner la plateforme (backend).

3.4 La zone d'archivage

Dans cette zone, qui pour la majorité du temps est déconnectée du réseau, se situent des serveurs de stockage à disques amovibles de plusieurs téraoctets. Le volume concerné est monté uniquement lors d'une sauvegarde ou lorsque de nouvelles données sont disponibles, minimisant le risque d'altérer les données sauvegardées.

3.5 La zone (ou cluster) de production

Cette zone est située sur un serveur physique distinct, fournissant des ressources de calculs nécessaires à l'exécution des algorithmes. Cette zone communique avec la zone de confiance via un pare-feu pour accéder aux données utilisées par les algorithmes.

4 Dépôt de données

Les données déposées sur la plateforme LOLA arrivent sous forme d'une archive chiffrée par le détenteur de celles-ci au moyen du chiffrement AES-256/RSA-2048. Cette clé publique de chiffrement asymétrique est partagée par les membres de la plateforme et sera renouvelée à chaque grande évolution de l'application de traitement (« Lola toolbox » est un outil est proposé à chaque partenaire afin de générer l'archive chiffrée avec la clé courante (`https://gitlab.inria.fr/lola/lola_toolbox/-/releases`)). Un service de transfert de fichiers a été mis en place sur la base du protocole Secure File Transfer Protocol (SFTP). Ce service est isolé du système d'exploitation dans un espace très restreint de la machine virtuelle. Les données transférées par ce service arrivent dans un répertoire dédié à cette tâche et sont déplacées automatiquement sur un autre serveur afin d'y être déchiffrées.
Afin de garantir un contrôle total sur la manipulation et l'accès aux données en clair, l'archive chiffrée est déplacée à un endroit inaccessible de l'extérieur, avant d'extraire les données de celles-ci au moyen de la clé privée de chiffrement pour être normalisées et importées dans une base de données. Les fichiers contenant les données en clair sont alors supprimées du système de fichier. La figure 3 suivante présente cette procédure, tandis que la figure 4 présente les zones impactées par le dépôt de données.

5 Circulation des données

La plateforme compte deux types de données sensibles :
— les logs pseudo-anonymisées fournies par les partenaires ;
— les identifiants des utilisateurs de LOLA (nom, prénom, adresse mail).

5.1 Les logs semi-anonymisées

5.1.1 Caractéristiques

Ces informations sont enregistrées dans une base de données TRAX-LRS sous le format standardisé xAPI. Chaque partenaire de la plateforme possède un accès à son propre jeu de données. Enfin une extraction partielle des données est accessible lors de l'exécution des algorithmes au sein de la plateforme. L'architecture de LOLA garantit une impossibilité de divulguer et d'accéder directement au contenu de la base de données au travers des scénarios d'exécution d'algorithmes.

5.1.2 Volumétrie

Ces informations représenteront plusieurs milliers de giga-octets à long terme. C'est pourquoi des serveurs de stockage réseau externes ou Network Attached Storage (NAS) sont utilisés pour l'archivage des sauvegardes.

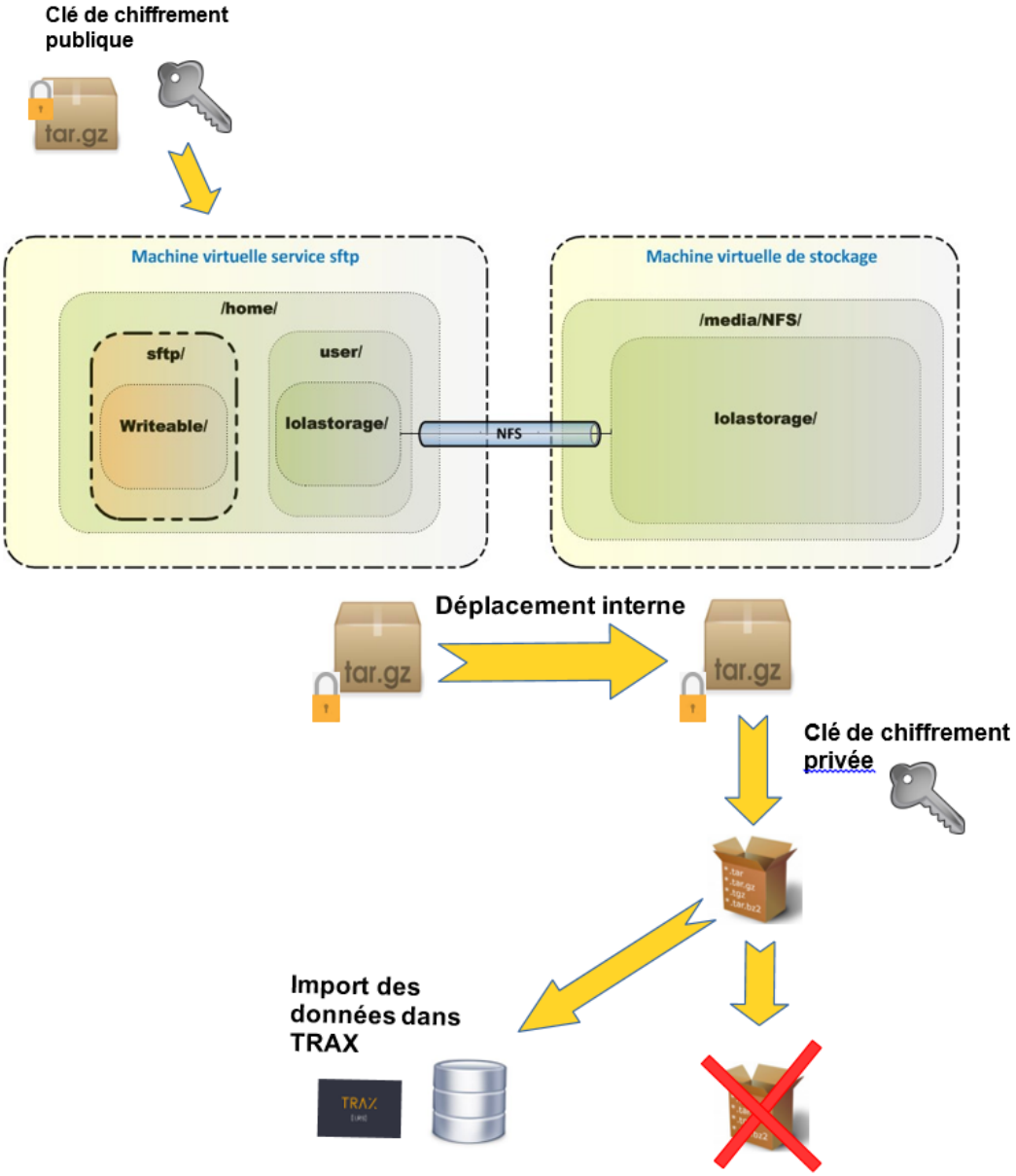

FIGURE 3 – Dépôt et transfert des données xAPI sur la plateforme LOLA

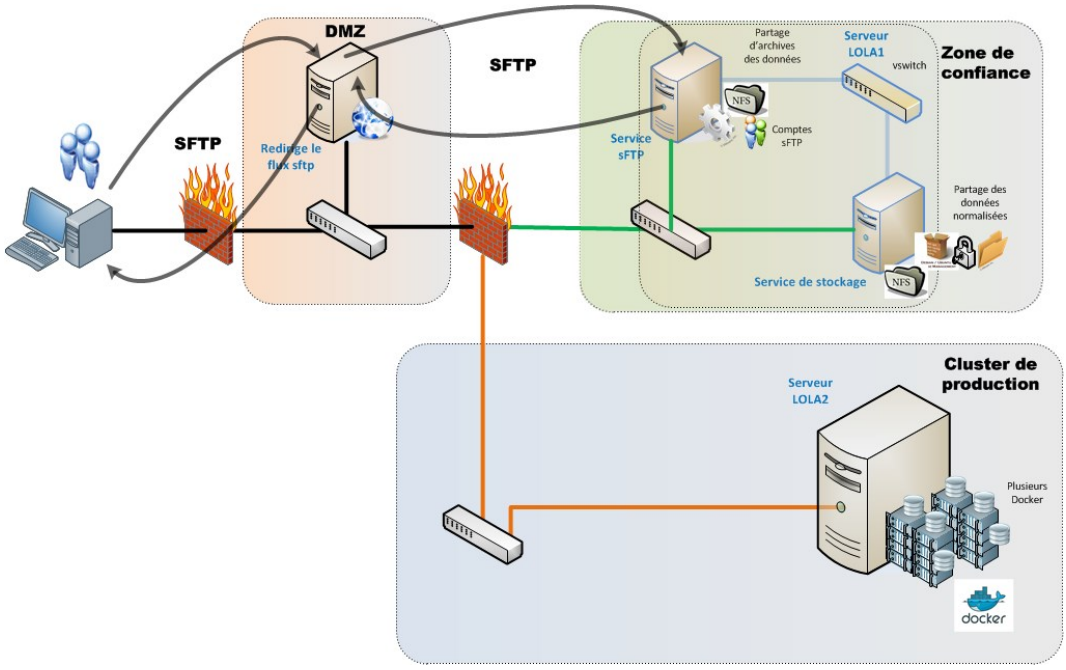

FIGURE 4 – Zones impliquées dans le dépôt et le transfert des données xAPI sur la plateforme LOLA

5.2 Les identifiants des utilisateurs

5.2.1 Caractéristiques

Tous les utilisateurs de la plate-forme sont identifiés par un login / mot de passe stockés dans une base de données dédiée à l'application web.

5.2.2 Volumétrie

Cela représente un très faible volume d'information à archiver (quelques centaines de mégaoctets au plus à long terme).

6 Plan de sauvegarde des données

Chaque mois les données sont extraites des bases et chiffrées ou exceptionnellement lors de l'intégration de nouvelles données.
Les bases de données opérationnelles ne sont pas chiffrées étant donné que leur contenu est confiné à deux serveurs (une machine virtuelle contenant le service de base de données et le serveur exécutant les scénarios).
De plus les clés de chiffrement associées aux archives sauvegardées sont également chiffrées avec une « phrase de passe » renouvelée régulièrement.