ITrust

IT Security Services

OSSIR

Reveelium

09 Février 2016

Reveelium

Smart predictive analytics

OSSIR

IKare©

My Business Monitoring in Real Time

Qui sommes nous ?

Equipe de Direction

Fondateur et Directeur Financier

H. P.

Ingénieur Supaéro,
Ancien PDG d'AIRBUS-ATR

Président Fondateur

Jean-Nicolas Piotrowski

Ancien **RSSI chez BNP Paribas**

CTO

J.L

Ingénieur ENSEEIHT,
15 ans d'expérience
en cybersécurité

VP

D. O.

Ingénieur Informatique - PhD en Management, HEC
20 ans d'exp .en business development High tech

Implantations

Paris, Toulouse (siège social)
San Francisco, 25 personnes

Positionnement

- ITRUST a été fondée en France en 2007 pour répondre aux besoins du marché en services de sécurité IT et a depuis :
  - Formé une équipe technique de haut niveau, rassemblant une combinaison unique d'expertises en cyber sécurité, Data-sciences et Intelligence artificielle
  - Développé une offre intégrant conseil, services managés et édition de logiciels innovants autour d'une vision orientée utilisateurs : « Keep Security Simple »
  - Développé des relation avec plus de 100 clients en France et à l'international

Chiffres clés

- 25 collaborateurs dont
  80% de docteurs & ingénieurs
- Plus de 100 clients publics et privés

Partenaires

SCC HISCOX AUSY ebrc Grant Thornton

Reveelium - le pitch

- Analyser les informations produites par le SI

  analyse basée sur les logs ou flux réseau.
- Observer et apprendre le fonctionnement du SI
- Créer des modèles comportementaux
- Détecter des anomalies, découvrir des schémas et des corrélations cachés sans nécessairement savoir ce que l'on cherche

Reveelium
Smart predictive analytics

Les fondations

- S’appuyer sur les éléments en place
  - Collecte des journaux souvent existante
  - SIEM : Agrégation et indexation des sources de données
- Technologie Big Data
  - Gestion diversité : des données, des entités et des relations
  - Scalabilité : augmentation du volume de données
- Analyse statistique et Machine Learning
  - Analytique & apprentissage
  - Détection d’anomalie

Réussite du Machine Learning

Premier acteur à utiliser cette technologie :

- Identifier le comportement des acheteurs
- Prévoir les tendances d'achat
- le plus visible : système de recommandations

Réussite la plus utilisée

- Contenu publicitaire en fonction des habitudes de navigation

Travaux plus utiles

- Observation des symptômes entrés dans Google
- Modélisation & prédiction de propagation d'épidémie

La « promesse » marketing

- Les solutions d'analyse comportementale remplacent les autres outils sécurité du SI
- Tous les types de données sont traités de manière transparente
- Toutes les anomalies sont automatiquement détectées

La réalité

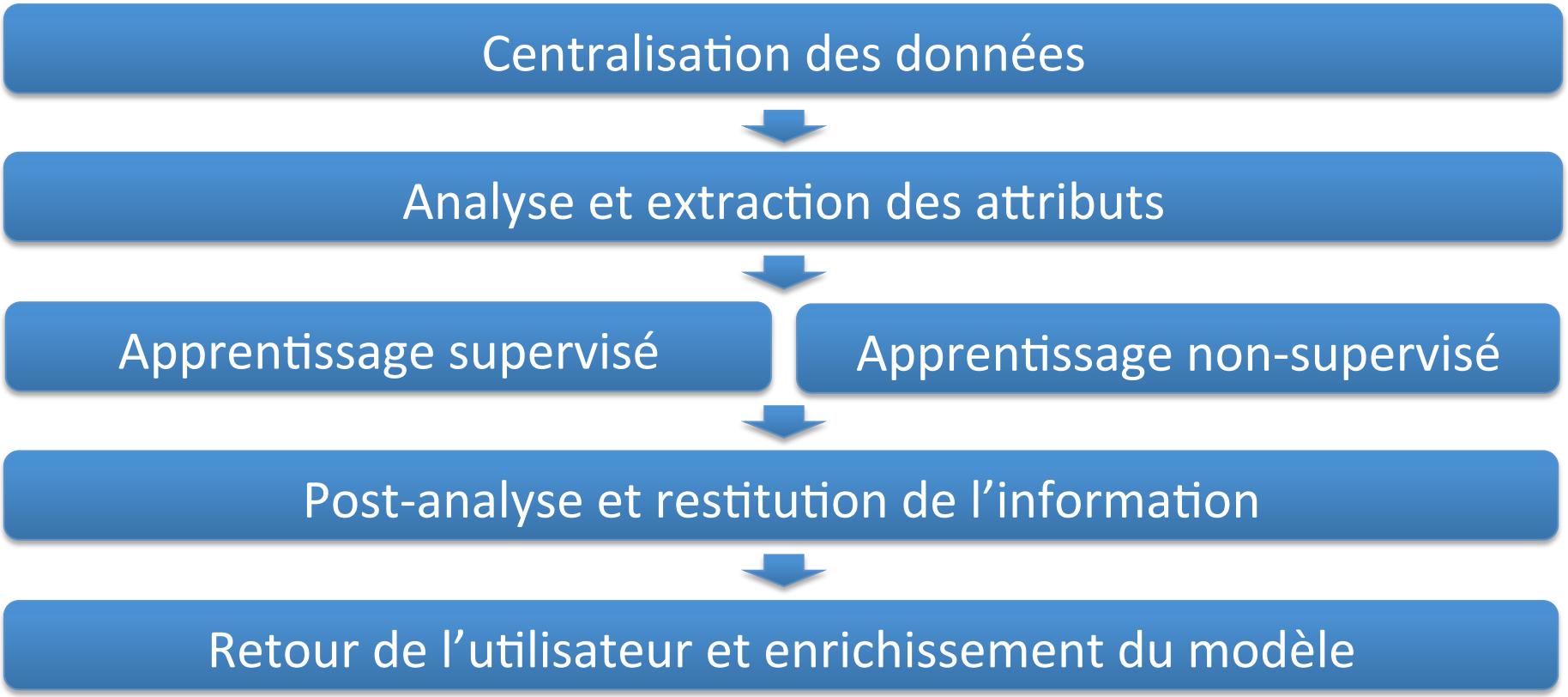
- Impossibilité d'un modèle unique détectant des anomalies dans n'importe quel type de données
- Les sources de données contraignent le type d'apprentissage
- Le modèle ne pourra être meilleur que la qualité des données en entrée

- Les algorithmes nécessitent forcément un apport externe
  - Compétences sécurité + data science
  - Données contextuelles externes

Freins liés à la sécurité

- Détections d'anomalies
  - La plupart des algorithmes sont équitables et équilibrés
- Comportement des utilisateurs
  - Les attaquants changent leur comportement pour éviter la détection
- Conséquences des erreurs
  - Ne pas laisser passer d'anomalies
  - Ne pas remonter trop de faux-positifs
- Jeux de données
  - Les données d'attaques non publiques

Méthodologie



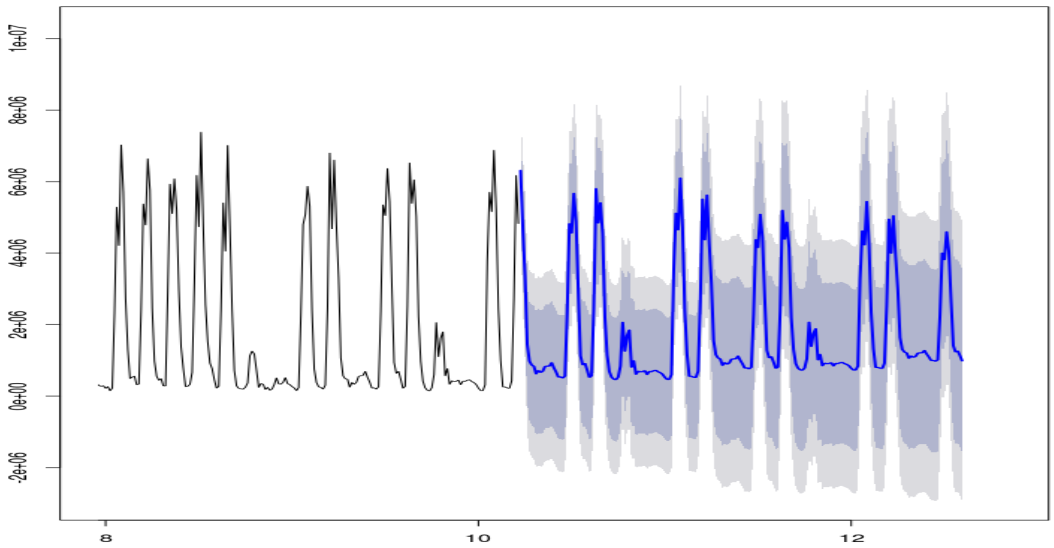
Revelium
Smart predictive analytics

Analyse temporelle

- Série temporelle

  Utilisation des statistiques et probabilités pour suivre l'évolution d'une caractéristique dans le temps

  - Volume, Fréquence, Unicité, Moyenne glissante
  - Comportement temporel, Tendances et Prédictions



Utilisation :

Utilisation intensive sur
- Les logs
- Les flux réseau

Analyse bande passante
Analyse volume de requêtes
Analyse de connexions utilisateurs
…

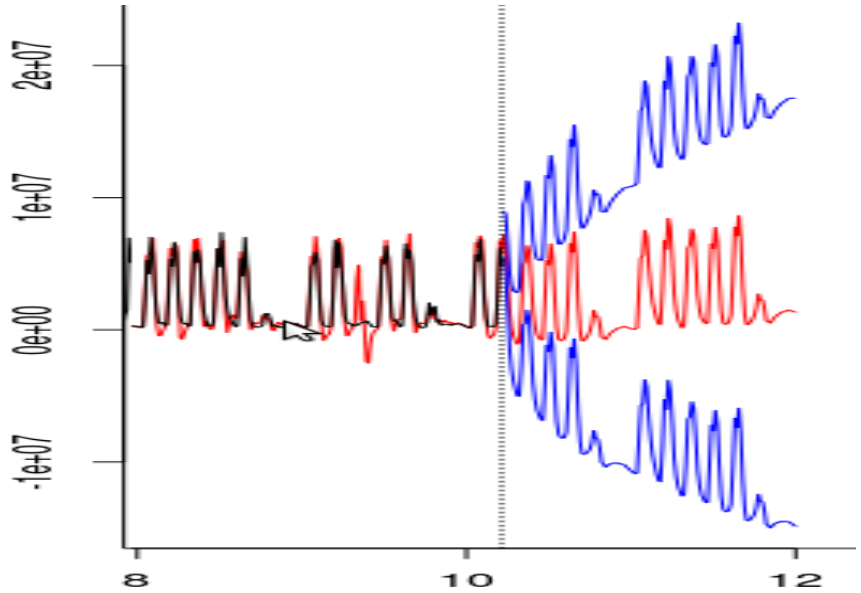
Revelium
Smart predictive analytics

Analyse temporelle

- Saisonnalité

  Identification de schémas se répétant à intervalles fixes

  - Décomposition des éléments temporels
  - Identification de saisonnalité multiples



Utilisation :

Utilisation sur
- Les logs
- Les flux réseau
- Des caractéristiques spécifiques

Identification d'usage anormaux
Identification d'évènements trop régulier
…

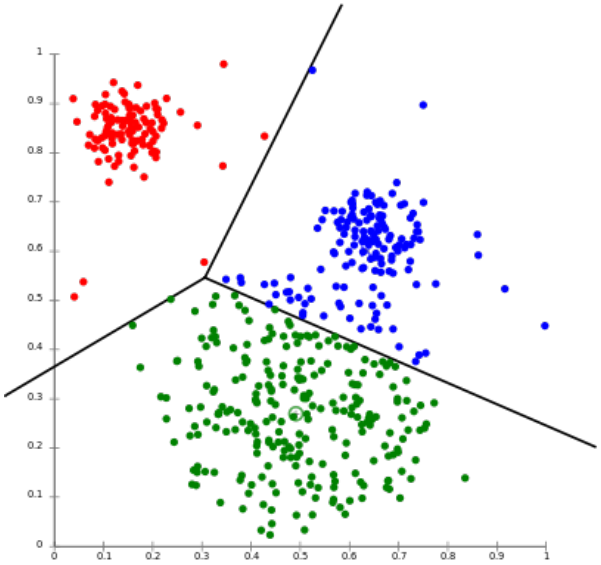
Revelium
Smart predictive analytics

Classification

- Classification / agrégation

  Identifier des similarités entre échantillons

  - Classer les échantillons dans des groupes
  - Rassembler les évènements par similarité



Utilisation :

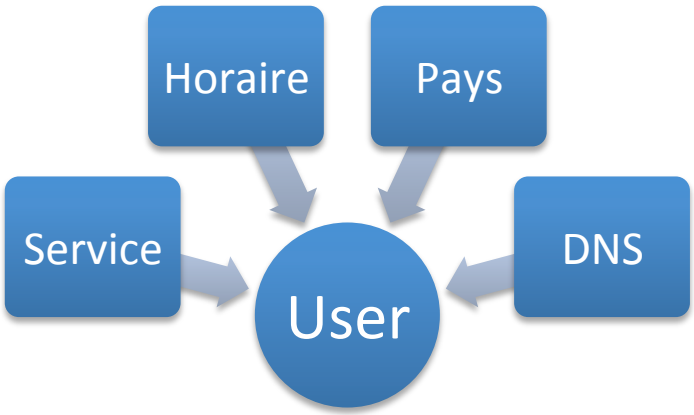
Utilisation sur
- Les métadonnées extraites des sources
- Caractéristiques spécifiques du modèle

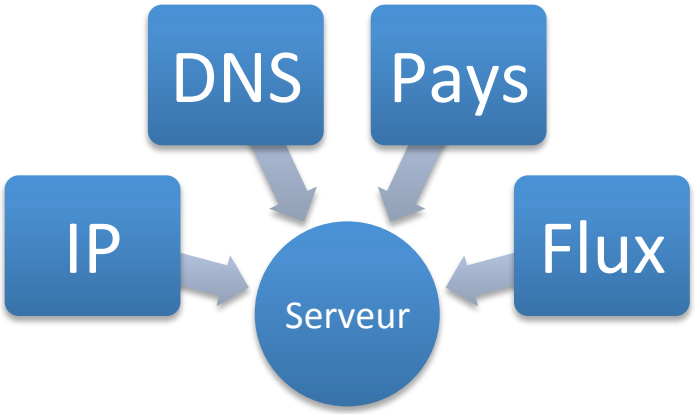
Identification d'échantillons anormaux
Classification des utilisateurs en fonction
- de leurs usages
- des services utilisés

…

Modèle comportementale

- Les composants logiques du réseau sont extraits pour construire les modèles
  - Modélisation des entités du réseau
  - Extraction des features d'une entité à partir des données
  - ML & analyse statistique utilisés pour apprendre les comportements
  - Détection de comportements déviants du modèle

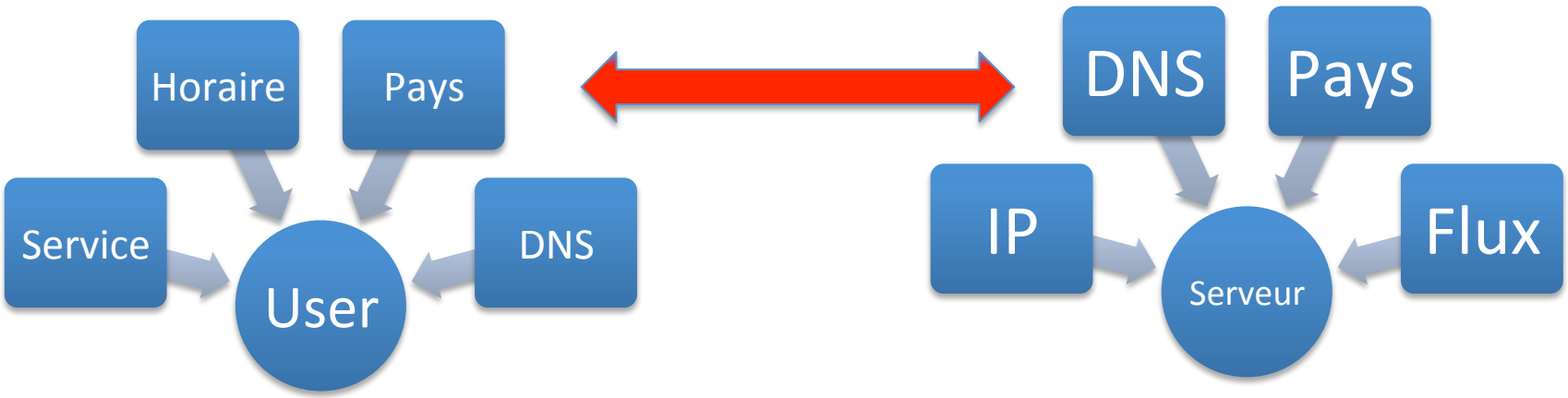




Revelium
Smart predictive analytics

Réseau d'entités

- Enrichissement de l'analyse dans un contexte réseau
  - Corrélation temporelle et évènementielle
  - Analyse de popularité
  - Création d'un réseau de relations (entre utilisateurs, serveurs, applications…) et analyse d'impact dans ce réseau



Revelium
Smart predictive analytics

Exemples d'anomalies

- Usurpation d'identité
  - Connexion utilisateur dans des zones géographiques différentes dans un laps de temps très court
- Extraction de données
  - Connexion d'un serveur vers des adresses IP et noms DNS inhabituels
- Mouvement latéral
  - Utilisateur se connectant à des serveurs inhabituels avec une fréquence élevée

De l'anomalie à l'alerte

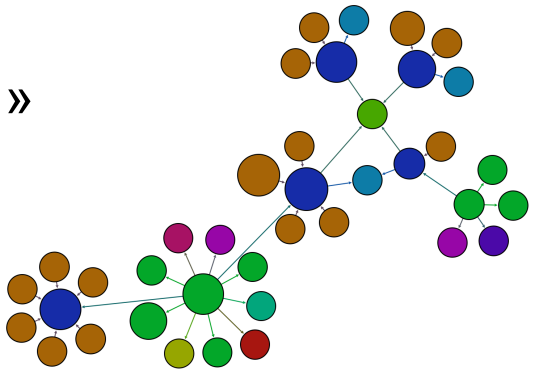
Apprentissage vs corrélation…

… pourquoi pas utiliser les 2 !

- Corrélation avec un système expert sécurité
- Utilisation de bases de connaissances (Threat Feed/ Intelligence) pour enrichir et contextualiser l'anomalie
- Partage des données dans une base de connaissance

Détection spécifique - DNS

- Machine Learning spécifique
  - Analyse des requêtes hors heures d'activité (night queries)
  - Analyse des requêtes vue pour la première fois
  - Analyse des techniques d'évasion de malware
    - DGA (Domaines Générés Aléatoirement)
    - DNS Tunneling
    - Fast Flux
- Relation & Théorie des graphs
  - Graph de connexion « Qui contacte qui ? »
  - Distribution géographique



Revelium
Smart predictive analytics

Analyse nom de domaine

DGA – Domaine généré aléatoirement

Cryptolocker domains

yrxtrwpncv.com
jowacrgnged.com
wbpbvtefxvh.com
znebqwgsqbkzu.com
iodgaudjyyafi.com
kydqgdnjacml.com
tjmlyxwfrf.com
ehincqzruzk.com
rulsxwnkallirdq.com
ogyinncagiiqx.com
kslttavhuczblq.com
uucaabmlzsp.com
nbiwbakdlchyowcdebanaqf.nu
ogcsgvdvpeokdbkk.com
psmdthlqxasoogq.in
pfrjquiuxiwnltyjy.su
vrsqnagcbtblimiperr.su
qgrgusynuwcdcvbfkykbggq.com
deehjyagmeqp.co

- Machine Learning spécifique
  - Analyse d'entropie / N-Gram
  - Longueur nom de domaine
  - Analyse géographique
  - Analyse syntaxique
- Machine Learning supervisé
  - Arbre de décision

Amélioration par l'expérimentation…

- Analyse de popularité

  Corrélation avec les habitudes des autres utilisateurs

- Détection de scan de port

  Identification de fréquence anormale de requêtes 1.0.168.192.in-addr.arpa

- Corrélation
  - NXDOMAIN
  - WHOIS

… et par l'échec

- Analyse de sous-domaine

  mauvaise idée : uuid VM cloud

⇒ Whitelist top 1M Alexa

  insuffisant mais nouvelle approche

  Analyse de similarité : identification phishing

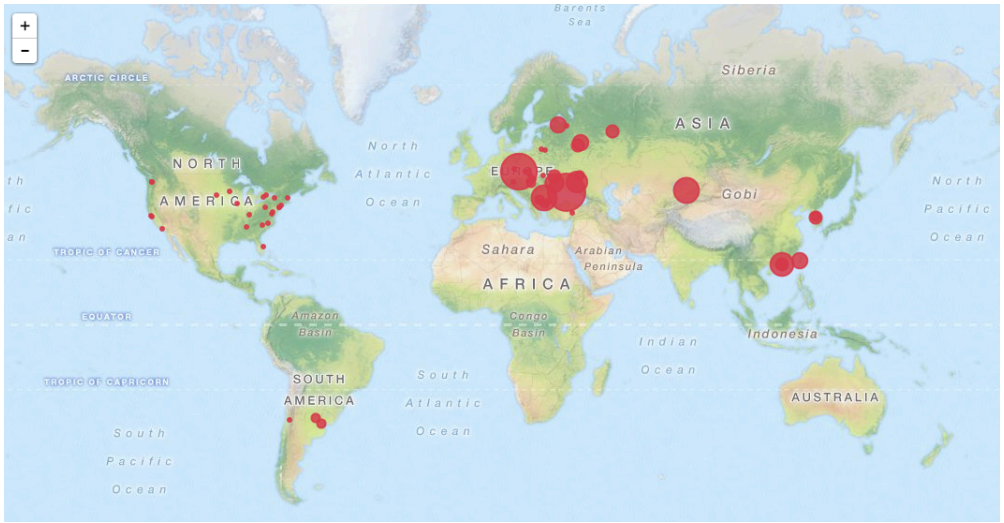
⇒ Variance : identification DNS tunneling

- Ajout des features dans les arbres de décision

Analyse TTL

Fast Flux – Analyse des réponses DNS

- Détection de domaine DNS avec de multiples adresses IP
- Analyse TTL : Bascule très rapide
- Analyse de répartition géographique



smartfoodsglutenfree.kz
(Zeus Tracker)

Registered : 2015-02-24
Période d'étude : 17/03 au 25/03
2278 adresses IP
420 AS
32 Pays

8 à 14 nouvelles IP toutes les 300 secondes

Revelium
Smart predictive analytics

Un mot sur l'architecture

Reveelium – plateforme de détection d'anomalie et de prévention

Défi technologique

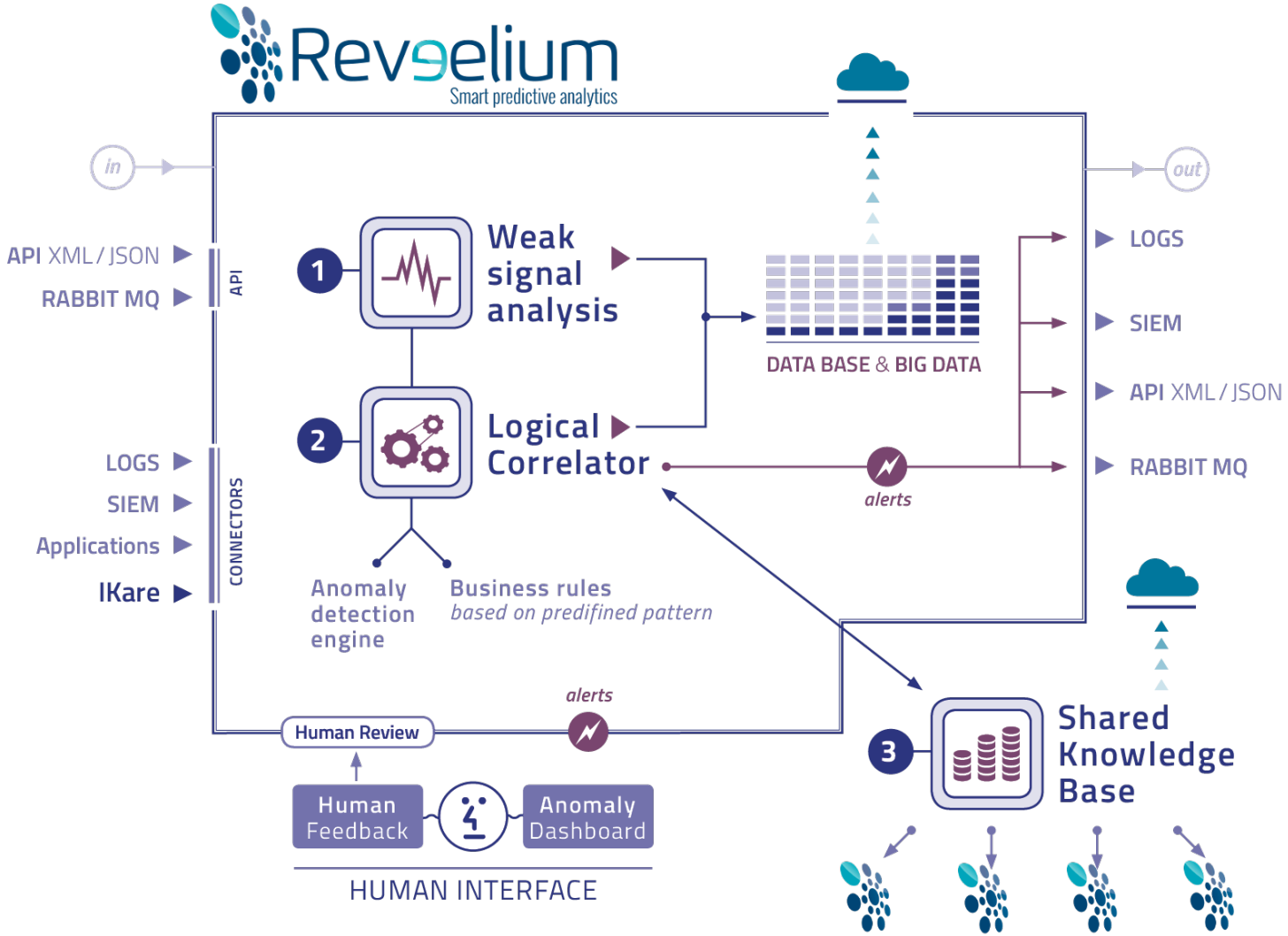
- Les solutions utilisent la même pile de traitement
  - Ingestion des données
  - Parsing
  - Indexation
  - Stockage
  - Processing analytique
- Chaque solution utilise sa pile propre
  - Perte de temps
  - Même données traitées plusieurs fois
- On refait les même erreurs que sur les SIEMs

Architecture générique

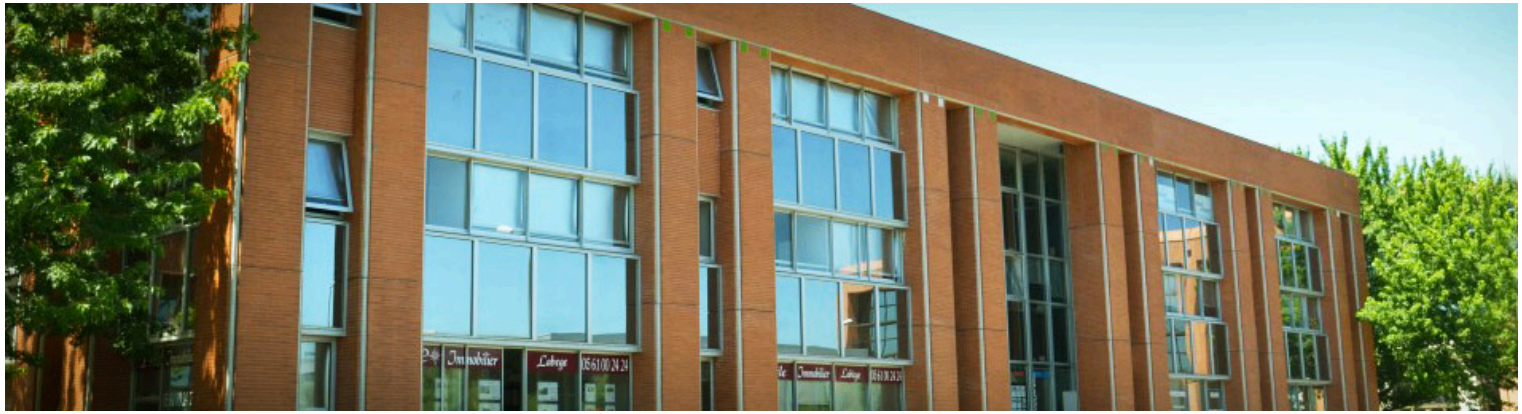
- Utilisation d'une architecture ouverte
- Basée sur des standards open source
  - Kafka (bus de communication)
  - Spark (Stream processor : indexation et calcul)
  - Elasticsearch (indexation / Base de connaissances)

- Pivot avec le reste du SI
  - Utilisation des données déjà traitées
  - Partage des anomalies et alertes
    - IDS / DNS Blacklist…
  - Base de connaissances requetable

Synthèse


Revelium
Smart predictive analytics
in
out
API XML / JSON
RABBIT MQ
API
LOGS
SIEM
Applications
IKare
CONNECTORS
1
Weak signal analysis
2
Logical Correlator
DATA BASE & BIG DATA
alerts
LOGS
SIEM
API XML / JSON
RABBIT MQ
Anomaly detection engine
Business rules
based on predifined pattern
alerts
Human Review
3
Shared Knowledge Base
Human Feedback
Anomaly Dashboard
HUMAN INTERFACE

Questions



ITrust - Siège Social
55 Avenue l'Occitane,
BP 67303
31673 Labège Cedex

+33 (0)5.67.34.67.80
contact@itrust.fr
www.itrust.fr
www.reveelium.com

Reveelium
Smart predictive analytics