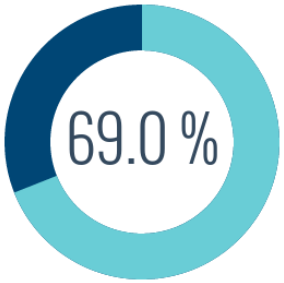
RéSIST
APT et analyse comportementale
15 Décembre 2015

Revelium
Smart predictive analytics

RéSIST

IKare©
My Business Monitoring in Real Time

Les défis de la sécurité



Des victimes découvrent l'attaque par une source externe

205

Nb de jours (en moyenne) pour détecter la présence d'un APT

70-90%

Des échantillons de malware
- sont uniques
- ciblés par entreprise

Target

TV5 Monde

RSA

Revelium
Smart predictive analytics

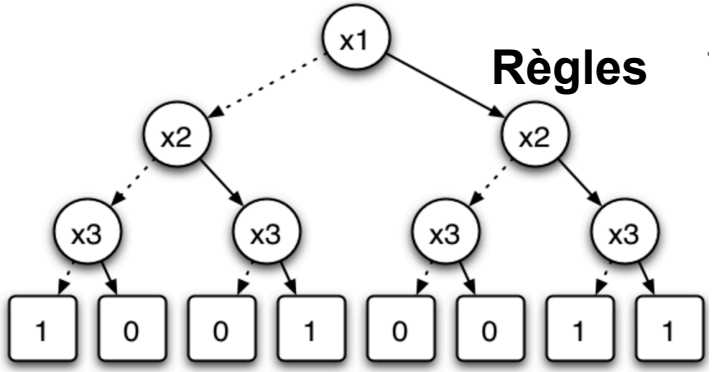
Les solutions actuelles sont impuissantes

Les nouvelles attaques surviennent
- trop vite
- trop souvent

La recherche de **signatures connues** ne suffit plus

Les règles de détections sont compliquées à mettre en œuvre.
Le temps et les compétences manquent.

S'appuient sur des schémas **d'attaques connues**



Les sandbox ne portent que sur la première phase d'une attaque.

Les malwares détectent et s'évadent des sandbox

Revelium
Smart predictive analytics

Au bon usage, le bon outil

Firewall

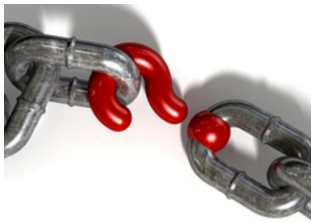
Protection périmétrique

IDS & antivirus

Protection contre les attaques connues (signatures)

SIEM

Collecte des traces et des informations
Génération d'alertes statiques et manuelles
Analyse manuelle



Chainon manquant:
Méthode d'auto-apprentissage dynamique pour répondre aux nouvelles attaques

Revelium
Smart predictive analytics

Alors on fait quoi ?

Revelium
Smart predictive analytics

La sécurité un problème de Big Data

- Evènements de plus en plus variés
  Nouvelles sources, entités, relations
- Volumes de données de plus en plus grand
  Besoin de scalabilité

- Besoin de découvrir des schémas et des corrélations cachées sans nécessairement savoir ce que l'on cherche (apprentissage machine, analyse statistique)

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

Freins liés à la sécurité

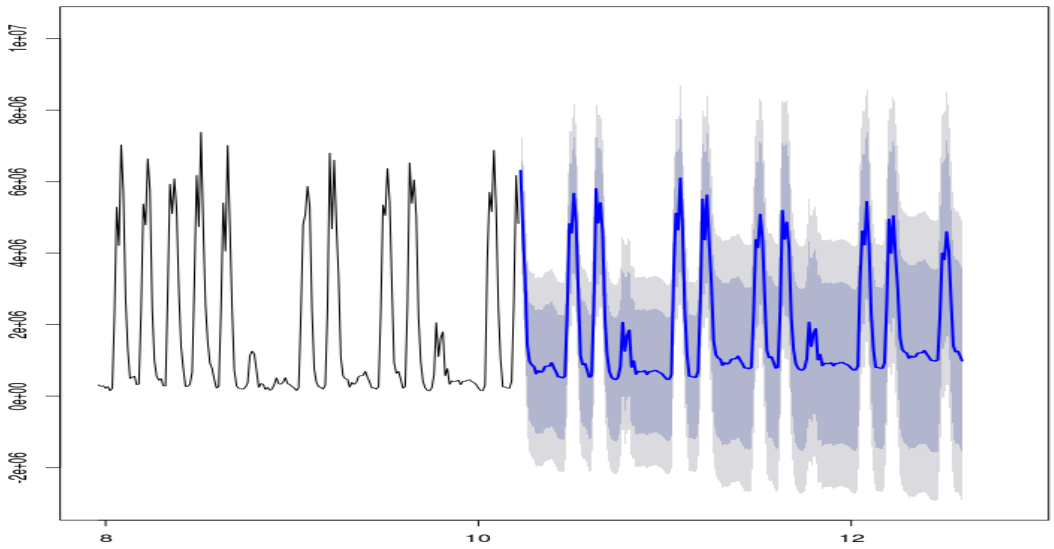
- Détections d'anomalies
  - La plupart des algorithmes sont équitables et équilibrés
- Comportement des utilisateurs
  - Les attaquants changent leur comportement pour éviter la détection
- Conséquences des erreurs
  - Ne pas laisser passer d'anomalies
  - Ne pas remonter trop de faux-positifs
- Jeux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

Modèles statistiques & Machine Learning

- 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
…

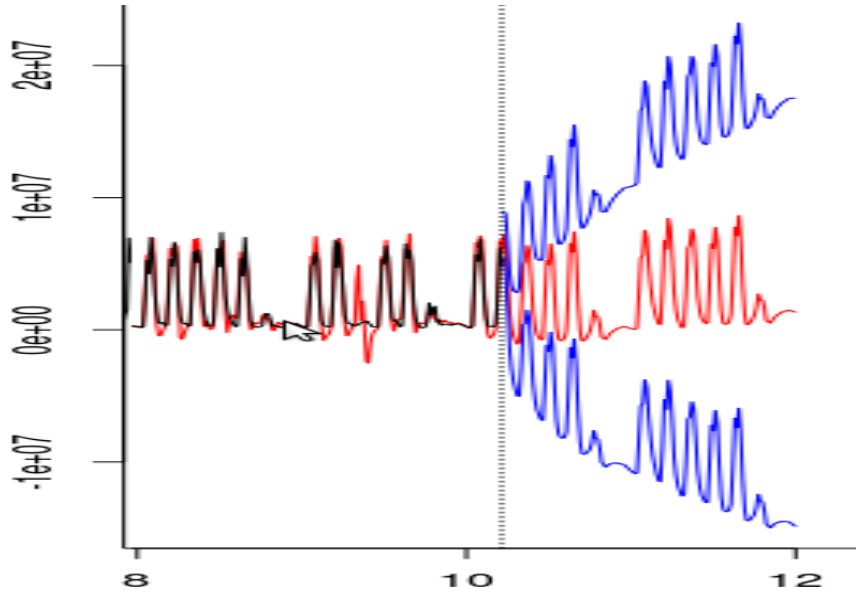
Revolelium
Smart predictive analytics

Modèles statistiques & Machine Learning

- 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
…

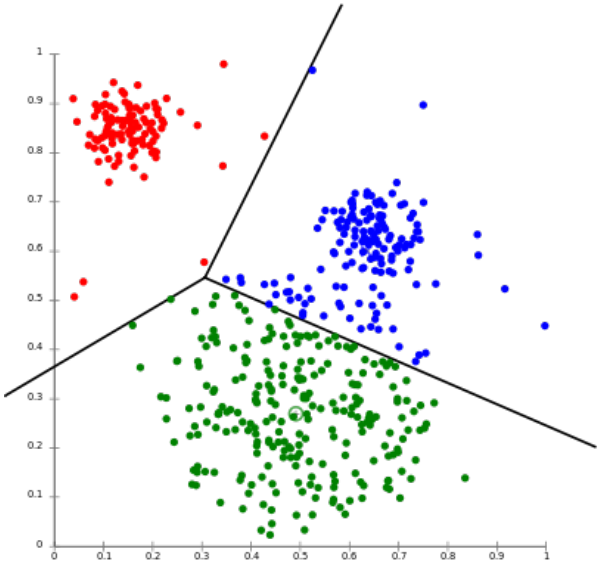
Revelium
Smart predictive analytics

Modèles statistiques & Machine Learning

- Classification / agrégation

  Identifier des similarités entre échantillons

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



Utilisation :

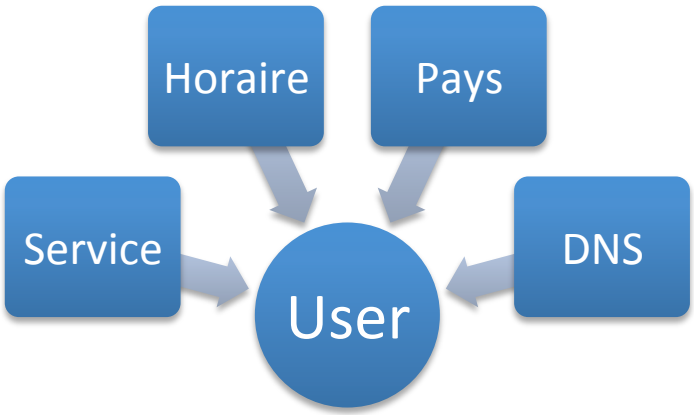
Utilisation sur
- Les métadonnées extraites des sources
- Caractéristiques spécifiques du modèle

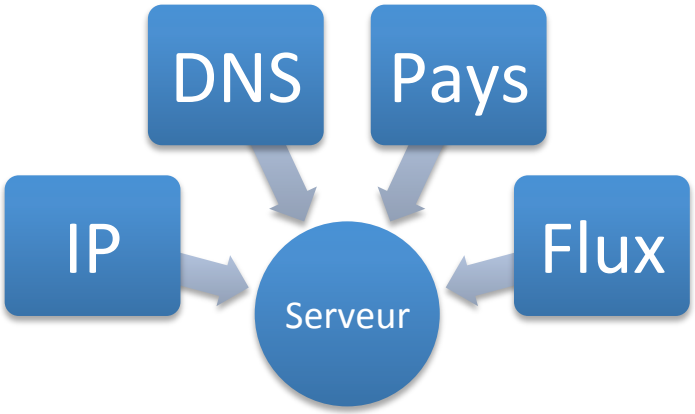
Identification d'échantillons anormaux
Classification des utilisateurs en fonction
- de leurs usages
- des services utilisés

…

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

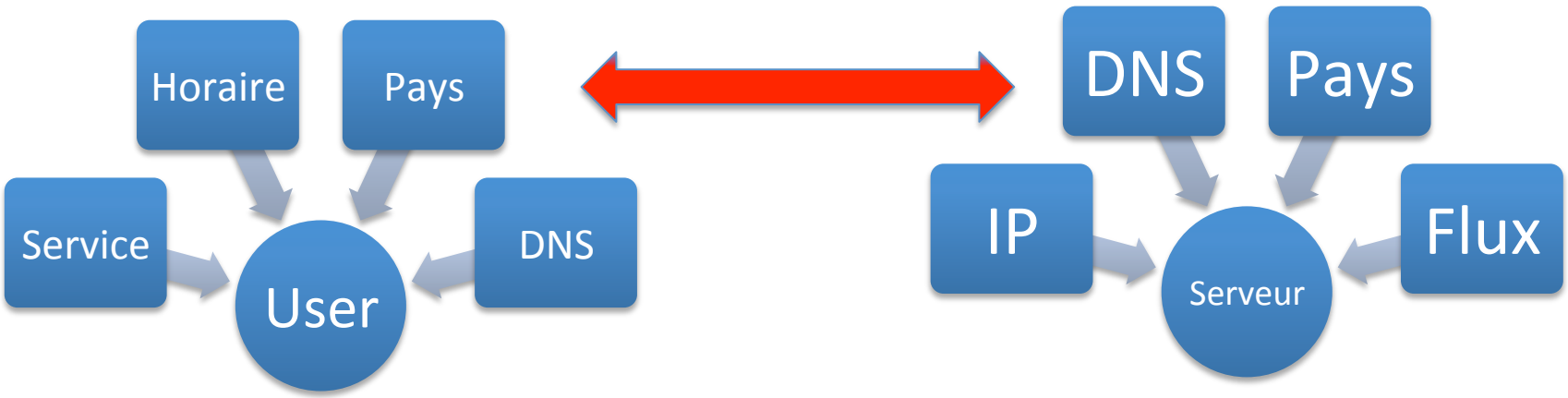




Revelium
Smart predictive analytics

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

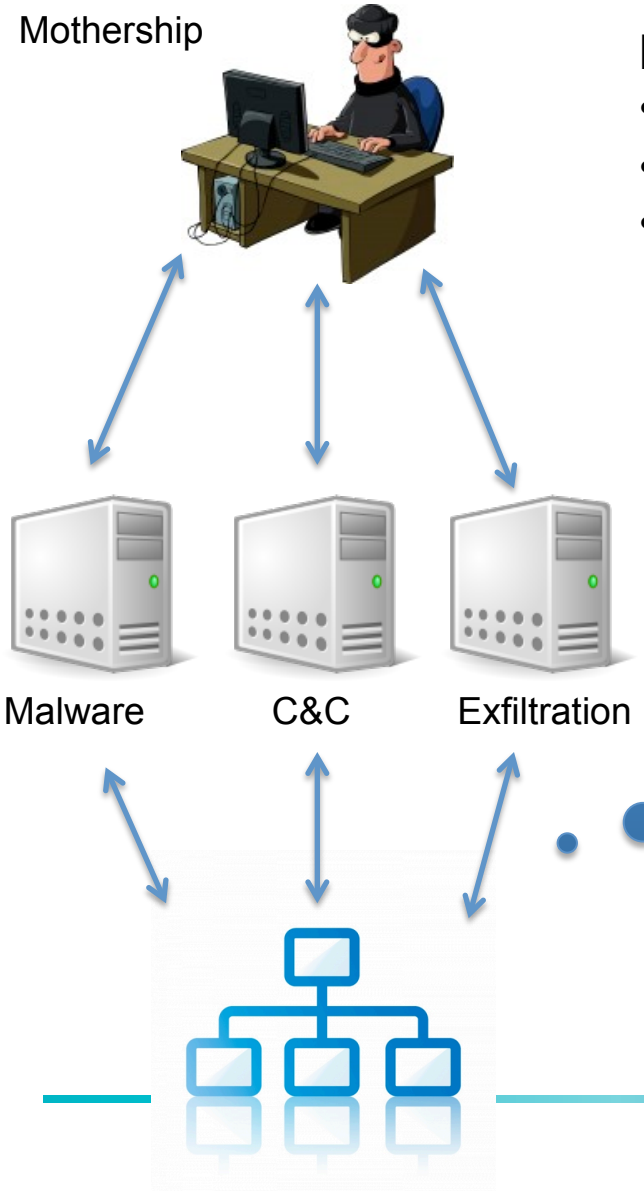
Cas pratique : détection d'APT

APT Kill Chain

- Reconnaissance
- Infiltration
  - Phishing
- Persistance
  - Installation de malware
  - Communication avec le serveur C&C
- Espionnage interne
  - cartographie et récolte d'informations
- Mouvement latéral
  - Propagation (utilisation de credentials)
  - Accès à l'information ciblé
- Exfiltration
  - exfiltration des données ciblées (But premier de l'attaque)

Reconnaissance → Infiltration → Persistance → Espionnage interne → Mouvement latéral → Exfiltration

DNS source essentielle



Le DNS est fortement utilisé lors des étapes de l'APT

- DNS scalable
- Robustesse en cas de démantèlement
- Evasion de blacklists IP

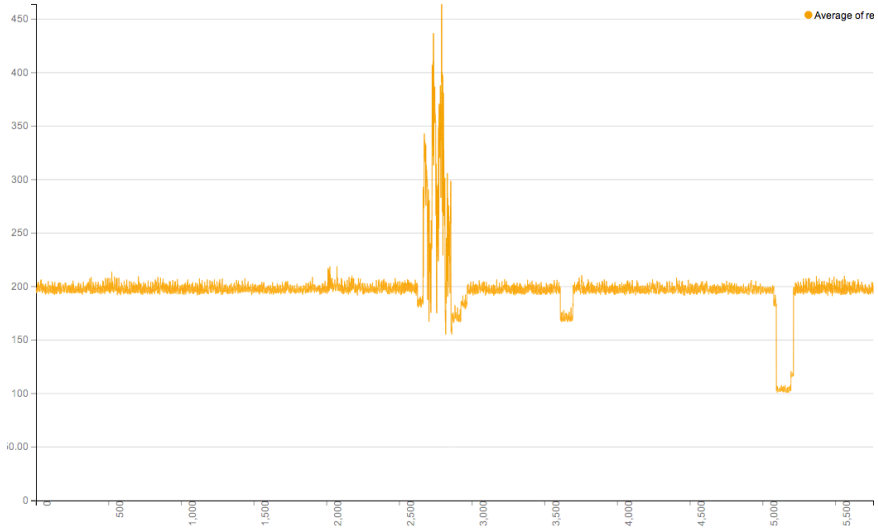
Superviser les requêtes DNS externes pour détecter des domaines suspects

Revelium
Smart predictive analytics

Analyse statistique

Profilage des données

- Jeu de données
  - 5,2 Millions de requêtes par jour
  - 50 000 domaines uniques
  - 2 000 sources IPs

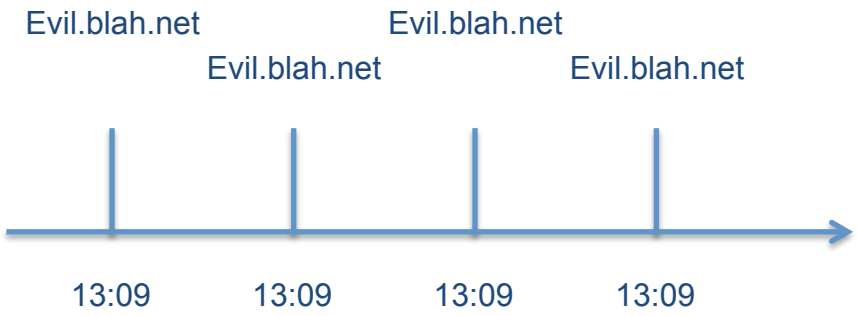


- Échantillons intéressants
  - Rafale de requêtes (burst)
  - Peu fréquentés : 10 000
  - Échantillon aberrants :
    - Domaines non enregistrés / résolus : 2000
    - Une source contacte 1000 domaines pointant sur la même IP

Revelium
Smart predictive analytics

Analyse statistique

- Saisonnalité
  - Identifier des évènements récurrents
  - Ex: Beaconing (notification d'installation réussie)
    - Périodicité parfaite -> activité automatique

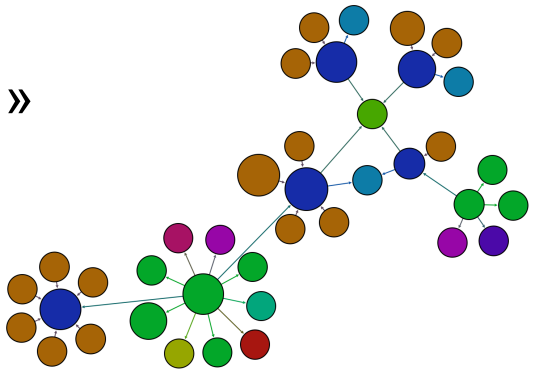


Revelium
Smart predictive analytics

Machine Learning

- Machine Learning spécifique au DNS
  - 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



Corrélation

- DNS Threat Intelligence
  - Réputation IP / DNS (~50 Bad Reputation)
  - Analyse WHOIS
  - Blacklist / Sinkhole (~500 domaines Blacklistés)
  - Alexa Top 1M (~7000 domaines hors du top)
  - Partage d'indice de compromission (IOC)

- DNS Threat Intelligence
  - Enrichissement de la base de données partagées

DGA – Analyse des noms de domaines

Cryptolocker domains

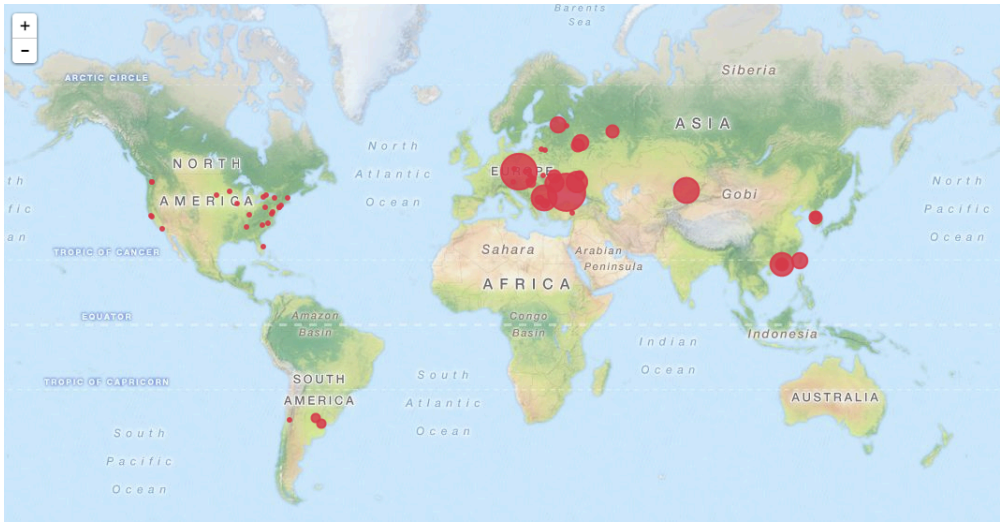
yrxtrwpncv.com
jowacrgnged.com
wbpbvtefxvh.com
znebqwgsqbkzu.com
iodgaudjyyafi.com
kydqgdnjacml.com
tjmlyxwfrf.com
ehincqzruzk.com
rulsxwnkallirdq.com
ogyinncagiiqx.com
kslttavhuczblq.com
uucaabmlzsp.com
nbiwbakdlchyowcdebanaqf.nu
ogcsgvdvpeokdbkk.com
psmdthlqxasoogq.in
pfrjquiuxiwnltyjy.su
vrsqnagcbtblimiperr.su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

- Machine Learning spécifique
  - Analyse d'entropie
  - Analyse de similarité
  - Analyse géographique
  - Analyse syntaxique
- Machine Learning supervisé
  - Arbre de décision

Technique d'évasion d'APT

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

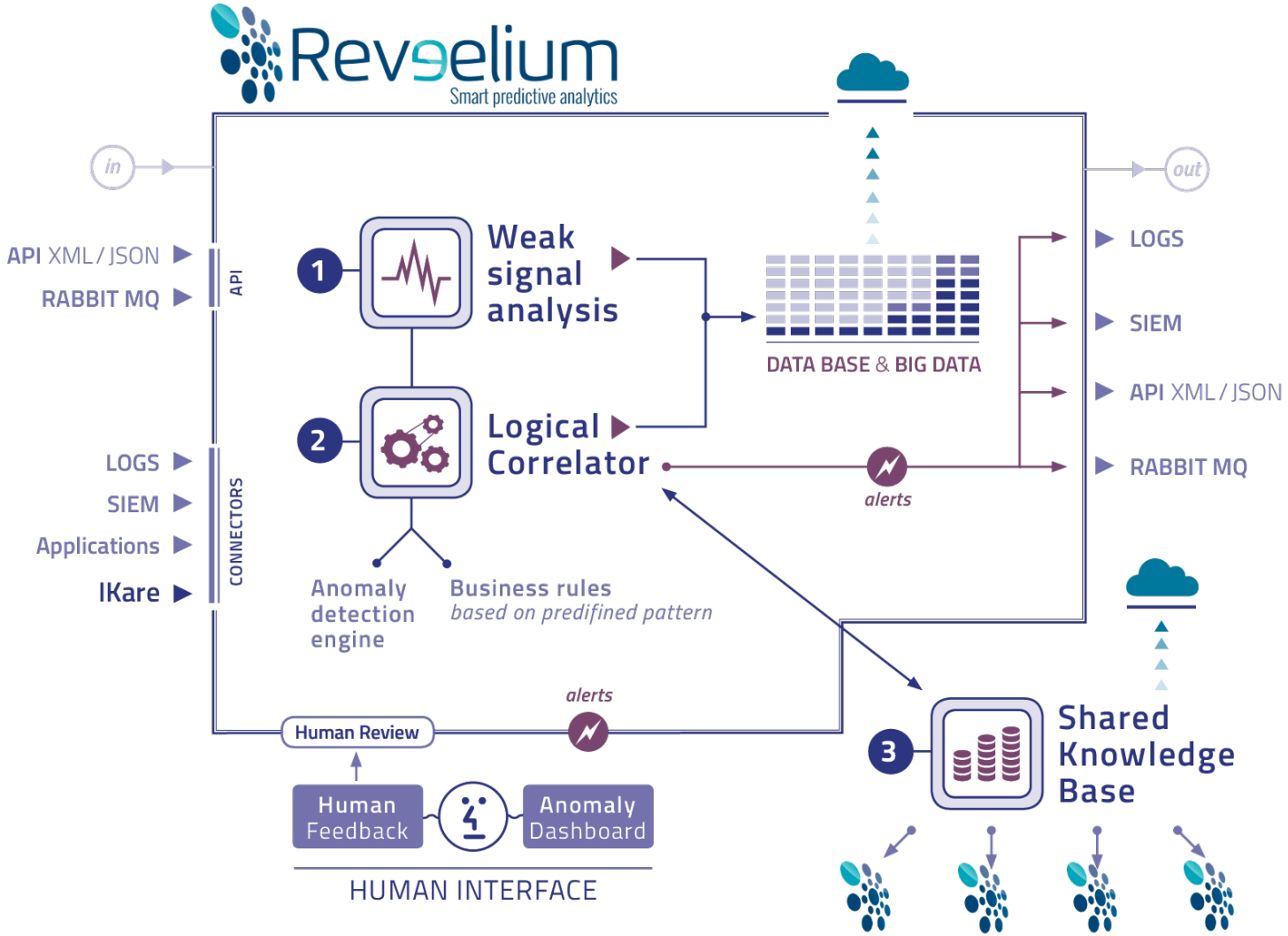
- Utilisation d'une architecture ouverte
- Basée sur des standards open source
  - Kafka (bus de communication)
  - Spark (Stream processor : indexation et calcul)
  - Elasticsearch (indexation / Base de connaissances)

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

Sources de données

- S’appuyer sur les sources existantes
  - Logs ou flux (data agnostic)
  - Data lake (Hadoop…)
  - Non intrusif : pas de sondes ou d’agents complémentaires
- Utiliser les points forts du SIEM
  - Agrégation et indexation des sources de données
  - Plugin ou intégré
- Fonctionnement agile
  - Streaming pseudo temps réel
  - Forensique

Synthèse

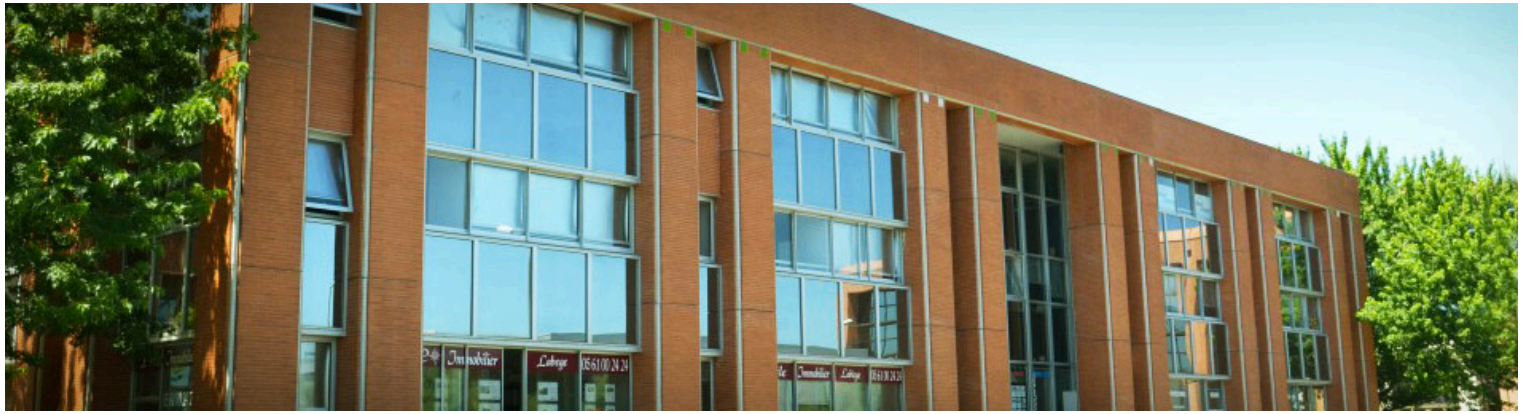

Revelium
Smart predictive analytics
in
out
API XML / JSON
RABBIT MQ
API
LOGS
SIEM
Applications
IKare
CONNECTORS
1
Weak signal analysis
2
Logical Correlator
DATA BASE & BIG DATA
alerts
LOGS
SIEM
API XML / JSON
RABBIT MQ
Anomaly detection engine
Business rules
based on predifined pattern
alerts
Human Review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

Capt'n Buzzword's Checklist

- ✓ APT
- ✓ Big Data
- ✓ Machine Learning
- ✓ Security Analytics
- ✓ Threat Intel
- ✓ Elasticsearch
- ✓ Docker

- ✗ Star wars

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