CryptoPanel

edycja #11

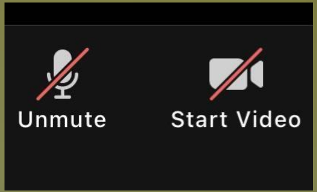
Lada moment zaczynamy...

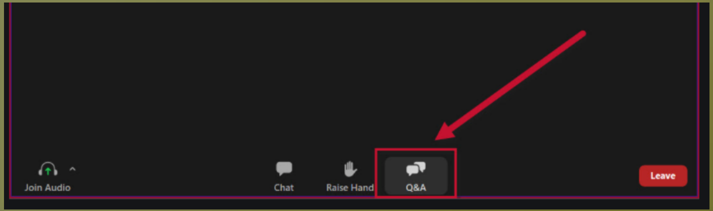
CryptoPanel

THALES

CLICO

CryptoPanel





CryptoPanel

edycja #11

Anonimizacja, tokenizacja i maskowanie – jak uczynić dane odpornymi na ujawnienie a jednocześnie móc z nich korzystać?

CryptoPanel

dziś dyskutują

Artur Holeczek

Partner Account Manager /
Product Manager
artur.holeczek@clico.pl
mob. +48 667 699 444

Piotr Majek

Security Specialist
piotr.majek@clico.pl
mob. +48 663 994 996

CryptoPanel

problem

co nas boli...

- Jesteśmy jednostką naukowo-badawczą
- Naszym zadaniem jest stworzenie ”zintegrowanej platformy badawczej (ZPB)”
- Platforma ma pozwalać przetwarzanie danych poprzez wykonywanie rozmaitych badań analitycznych (nie tylko statystyka!)
- Dawcami (donorami) danych będą różne instytucje państwowe.
- Dane dostarczane przez donorów mogą być danymi wrażliwymi (np. dane o obywatelach)
- Platforma musi zapewnić pełną kontrolę i transparentność nad tym jakie dane przekazuje donor
- Platforma musi zapewnić na potrzeby donora:
  - Szyfrowanie,
  - Anonimizację,
  - Tokenizację (w tym jednokierunkową) danych.

- ...jak zapewnić bezpieczeństwo danych gdy są one przekazywane do platformy analitycznej?
- …jak zapewnić tokernizację, anonimizację danych wykonywaną i kontrolowaną przez donorów?
- …czy możemy zapewnić utrzymanie formatu danych (np. nr PESEL, IMEI, itp.)?
- …jak zapewnić brak możliwość „odtokenizowania danych”?
- ...jak zapewnić koherencję tych samych danych pochodzącą od różnych donorów/repozytoriów (np. nazwy miast, nazwiska czy numery NIP, itp)
- …*pseudonimizacja* – czym różni się od tokenizacji, anonimizacji i szyfrowania?

Dla ustalenia uwagi: pseudo- tokenizacja a szyfrowanie?

Pseudonimizacja jest odwracalnym procesem, który jest przeprowadzany w celu ograniczenia możliwości identyfikacji danej osoby. W ra
za zadanie wyszczególnić dane, które pozwalają zidentyfikować daną osobą fizyczną, a następnie zastąpić je innymi danymi np. pseudo
procesowi pseudonimizacji dalej podleg

Metody pseudonimizacji danych [edytuj | edytuj kod]

Administratorzy danych stosują różne metody pseudonimizacji danych, jednak najczęściej używane to:

-tokenizacja- polega na zmianie danych na losowo wygenerowany ciąg liczbowy,

-skracanie- polega na skróceniu danych pozwalających na identyfikację danej osoby fizycznej,

-szyfrowanie tajnym kluczem- polega na szyfrowaniu danych za pomocą klucza szyfrującego, który w przyszłości może być użyty do odszyfrowania danych.[3]

Pseudonimizacja a anonimizacja [edytuj | edytuj kod]

Anonimizacja polega na usunięciu wszelkich danych umożliwiających identyfikację danej osoby, natomiast pseudonimizacja polega na przetworzeniu danych, tak aby uniemożliwić ich powiązanie z daną osobą. Anonimizacja jest procesem trwałym i nieodwracalnym, natomiast pseudonimizacja jest procesem odwracalnym[4].

Tokenization

From Wikipedia, the free encyclopedia

Tokenization may refer to:

- Tokenization (lexical analysis) in language processing
- Tokenization (data security) in the field of data security
- Word segmentation
- Tokenism of minorities.

Tokenization (data security)

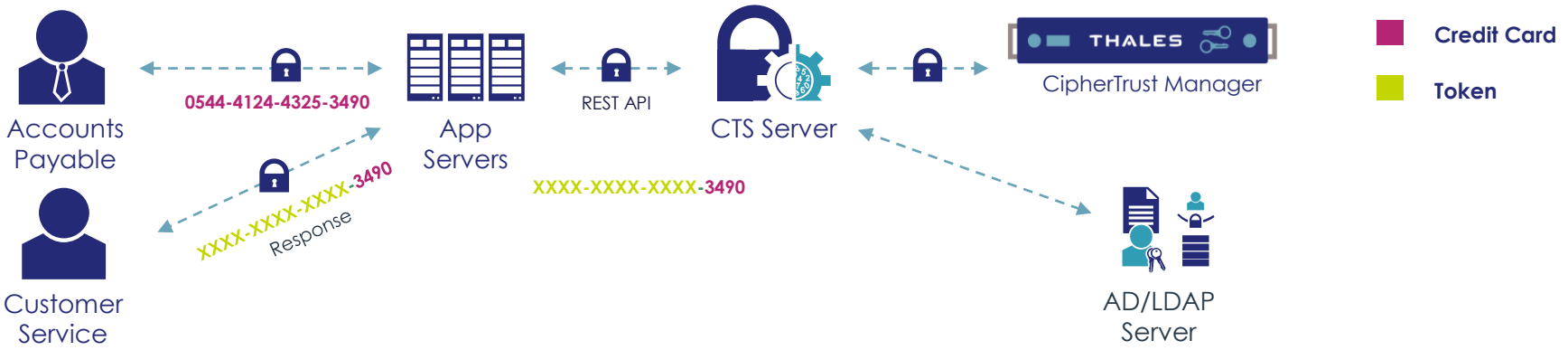
From Wikipedia, the free encyclopedia

Not to be confused with tokenization (lexical analysis).

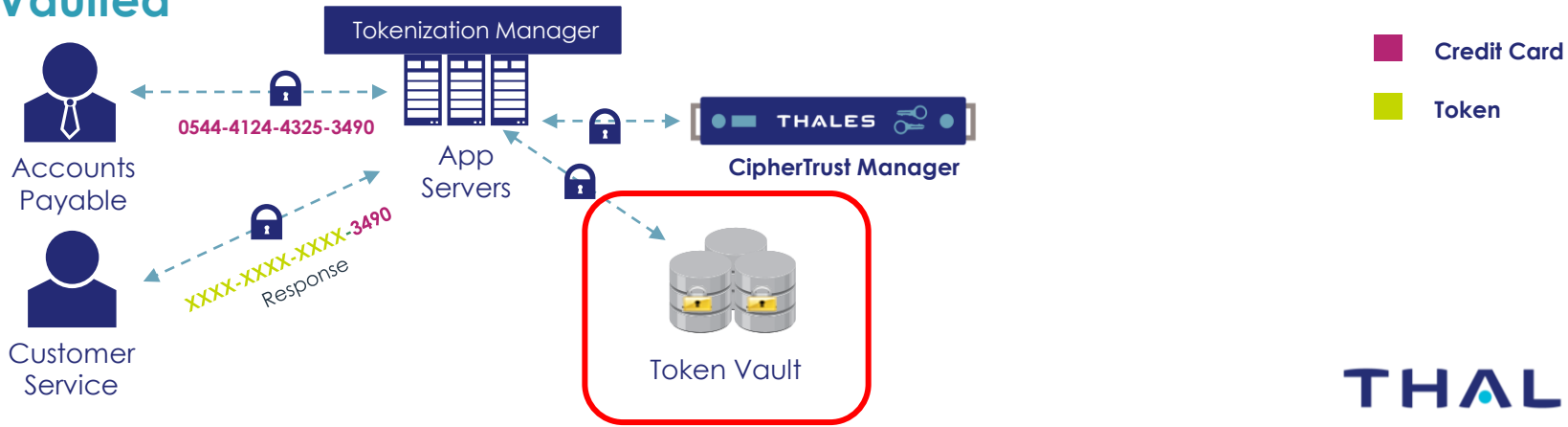
Tokenization, when applied to data security, is the process of substituting a sensitive data element with a non-sensitive equivalent, referred to as a token, that has no extrinsic or exploitable meaning or value. The token is a reference (i.e. identifier) that maps back to the sensitive data through a tokenization system. The mapping from original data to a token uses methods that render tokens infeasible to reverse in the absence of the tokenization system, for example using tokens created from random numbers.[3] The tokenization system must be secured and validated using security best practices[4] applicable to sensitive data protection, secure storage, audit, authentication and authorization. The tokenization system provides data processing applications with the authority and interfaces to request tokens, or detokenize back to sensitive data.

Tokenizacja bezstanowa i stanowa

Vaultless



Vaulted



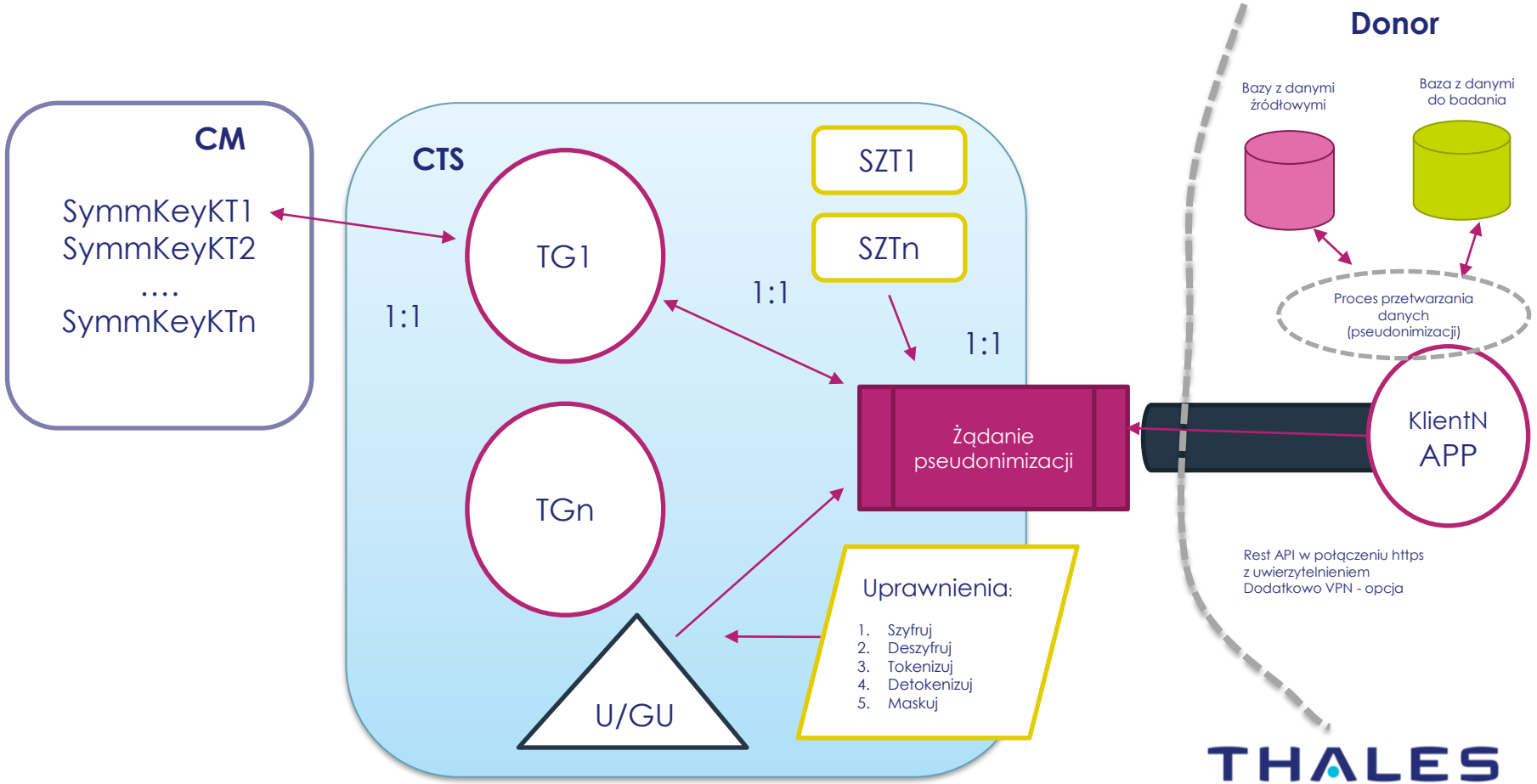
THALES

CryptoPanel

rozwiązanie

Rozwiązanie – CM z CTS i ewentualnie CDAP i wszystko jest możliwe

Przebieg procesu pseudonimizacji na platformie badawczej



Przebieg akwizycji danych do ZPB

1. *Dla przeprowadzenia pobrania danych do badania (B) tworzymy:*
   - *Klucz symetryczny w CM (AES256) - posłuży on do budowy wszystkich tokenów w ramach TG dla metod z zachowaniem formatu (FPE) i jako ziarno generatora PRNG w metodzie RANDOM*
   - *Grupę tokenizacji (TG) w CTS, którą kojarzymy z kluczem symetrycznym w CM*
   - *Użytkownika w CTS nadając mu odpowiednie uprawnienia w CTS (np. szyfruj, tokenizuj). Użytkownikiem tym będzie posługiwać się aplikacja klienta (KlientAPP) dla uwierzytelnienia w odwołaniach REST API*
2. *Wszyscy donorzy danych w danym badaniu muszą pseudonimizaować dane w ramach tej samej TG dla zachowania koherencji danych*
3. *Tworzymy szablony przekształcania danych (SZT). Szablony mogą podlegać wykorzystaniu dla wielu klientów. SZT definiują proces przekształcania danych i upraszczają wywołania REST API*

KlientApp

1. **KlientAPP nie zawiera żadnych poświadczeń ani kluczy szyfrujących. Aplikacja jest łatwa w audycie po stronie Donora (skrypt, java, itp.)**
2. **Poświadczenia użytkownika (użytkownik w CTS)**
3. **Aplikacja dokonuje uwierzytelnienia w CTS dla sesji Rest API (uzyskanie tokena sesji)**
4. **Aplikacja otwiera zbiór danych wejściowych**
5. **W pętli aplikacja dokonuje:**
    1. Parsowania zbioru danych wejściowych
    2. Dla danych wrażliwych wykonuje zapytanie POST do CTS, uzyskuje token(y) i zapisuje do zbioru wynikowego
    3. Dla danych niewrażliwych zapisu danej do zbioru wynikowego
6. **Po przetworzeniu zbioru Aplikacja szyfruje i przesyła zbiór wynikowy do platformy badawczej (ZPB)**

THALES

Bezpieczeństwo ZPB

1. Zabezpieczenie przed przechwyceniem danych wrażliwych:

1. Dane wrażliwe tokenizowane są w oddzielnych zapytaniach (atomowo) – brak korelacji z innymi danymi z rekordów źródłowych (np. w kolejnych zapytaniach przetwarzane są tylko nr PESEL)
2. Wywołania RestAPI z uprzednim uwierzytelnieniem (token sesji)
3. Zapytania RestAPI są realizowane w połączeniu szyfrowanym HTTPS do serwera CTS
4. Możliwość zestawienia dodatkowego połączenia VPN miedzy infrastrukturą Donora, a bramą ZPB
5. Wykorzystanie tylko funkcji skrótu (SHA2 lub SHA3) do przetwarzania danych w KlientAPP – brak zapytań RestAPI do CTS
6. Wyłączenie/skasowanie użytkownika w CTS związanego z aplikacją KlientAPP (brak możliwości uwierzytelnienia w RestAPI)

2. Zabezpieczenie przed możliwości dotarcia do danych pierwotnych:

1. Szablon tokenizacji (SZT) z cechą „nieodwracalne" (*irreversible*)
2. Przydzielenie użytkownikowi w CTS tylko prawa „tokenizuj"
3. Po wykonaniu akwizycji danych usunięcie TG nie pozwala na wykonanie operacji de- i tokenizacji
4. Po wykonaniu akwizycji usuniecie klucza (SymmKeyN) skojarzonego z TG w CTS co uniemożliwia powrót do danych wejściowych.

THALES

Przykłady wywołań

POST https://primaryvts.mylab.local/vts/rest/v2.0/tokenize Send

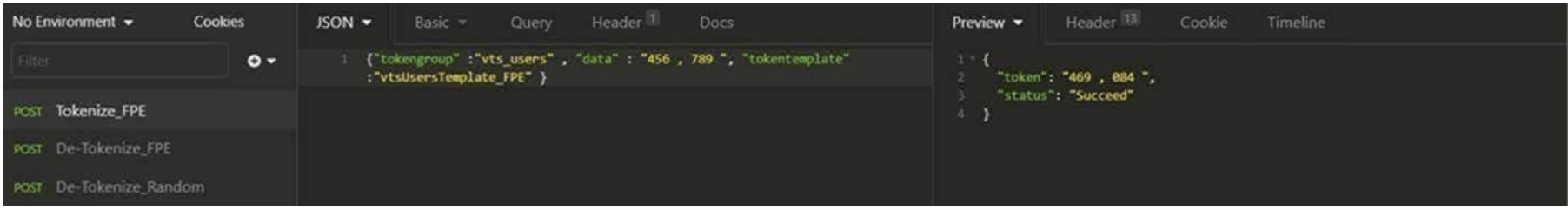
200 OK 0 ms 62 B

JSON Basic Query Header 1 Docs

```
1  {"tokengroup" :"vts_users" , "data" : " i93849382094023ue09u2ej9ru0239",
   "tokentemplate" :"vtsUsersTemplate_FPE" }
```

Preview Header 12 Cookie Timeline

```
1 {
2   "token": " pU3X7hOYK5GKor1fXf8UxV06iXK0hb",
3   "status": "Succeed"
4 }
```



POST https://primaryvts.mylab.local/vts/rest/v2.0/tokenize Send

200 OK 16 ms 43 B

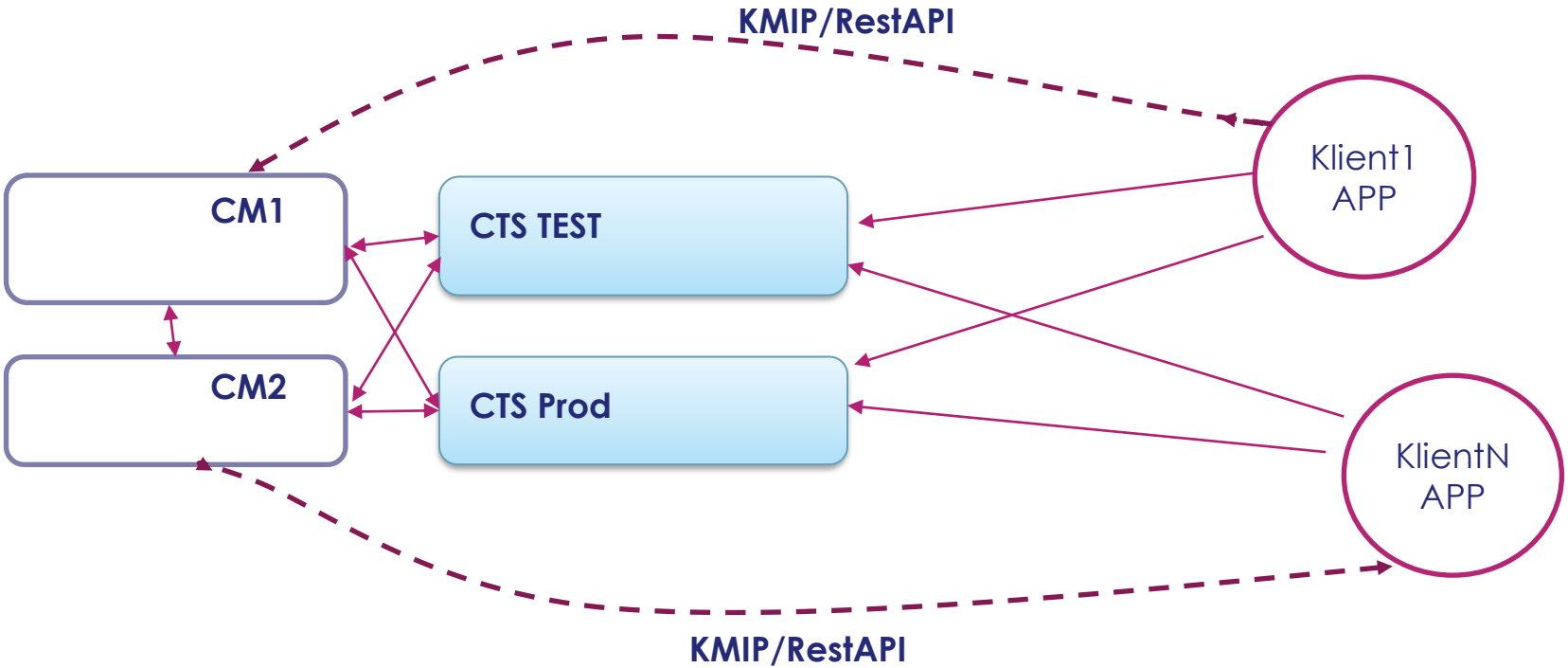
JSON Basic Query Header 1 Docs

```
1  {"tokengroup" :"vts_users" , "data" : "St Barbary 2", "tokentemplate"
   :"vtsUsersTemplate_FPE" }
```

Preview Header 12 Cookie Timeline

```
1 {
2   "token": "aH 1Ft4LXy Q",
3   "status": "Succeed"
4 }
```

Architektura – propozycja rozwiązania



THALES

„Nauczki“, czyli *lessons learned #1*

- **Czy jest sposób na uzyskanie statystyk pokazujących adresy IP (lub nazwy hostów) maszyn wykonujących wywołania kryptograficzne REST do serwera CT-VL? Pulpit nawigacyjny wyświetla się tylko według daty i można go filtrować, aby pokazać użytkowników, ale nie maszyny źródłowe…**

Subject: RE: CT-VL how to track REST call source machines?

Aha! I noticed with log level set to INFO that the tokenization.log had enough info:

May 20 10:39:08 ctshenry journal: ctshenry.vds.net nginx: 2022/05/20 10:39:08 [info] 19980#0: *1 ngx_http_auth_request_handler() : processing basic auth case while sending to client, client: 192.168.2.67, server: vts-0, request: "POST /vts/crypto/v1/encrypt HTTP /1.1", host: "192.168.2.101"
May 20 10:39:08 ctshenry journal: ctshenry.vds.net nginx: 2022/05/20 10:39:08 [info] 19980#0: *1 ngx_http_auth_request_handler(): jwt accepted, will be entered into JWT cache while sending to client, **client: 192.168.2.67, server: vts-0, request: "POST /vts/crypto/v1/encrypt HTTP/1.1", host: "192.168.2.101"**
May 20 10:39:08 ctshenry journal: ctshenry.vds.net nginx: 2022/05/20 10:39:08 [info] 19980#0: *1 ngx_http_auth_request_handler(): processing response from authentication microservice, user 'superuser' entered into JWT cache while sending to client, **client: 192.168.2.67**, server: vts-0, request: "**POST /vts/crypto/v1/encrypt HTTP/1.1**", host: "192.168.2.101"

Tego raczej nie znajdziecie (chyba że w dokumentacji)…

A co z wydajnością?

- 10000 zapytań Rest API per 1 CTS

Co jest szybsze?

- Szyfrowanie
- Tokenizacja VT-L VT
- Maskowanie
- …

Ograniczenia

- Minimalny długość wyrażania do tokenizacji (2)
- Znaki diakrytyczne?

Minimum input characters with keepleft or keepright

If you configure using keepleft or keepright, note the following minimum numbers of input characters, according to tokenization format:

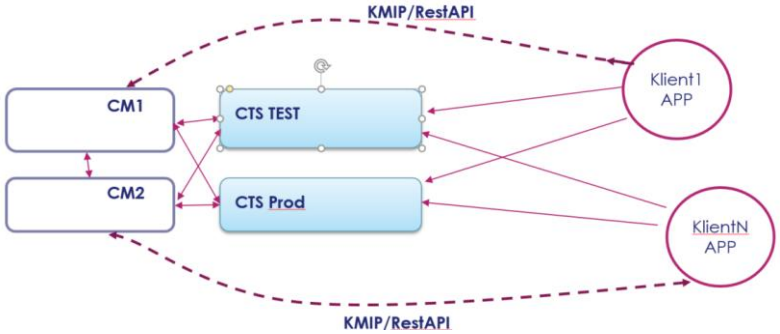
- For FPE: 2 characters plus the number of keepleft or keepright characters
- For RANDOM Numeric: 9 characters plus the number of keepleft or keepright characters
- For RANDOM Alphaumeric: 5 characters plus the number of keepleft or keepright characters

CryptoPanel

podsumowanie

Podsumowanie, czyli co potrzeba…

Licencje



- 2x CM (k170v)
- 2x licencja CTS
- 2x licencja CADP
- Opcja: 2x Flex Connector (KMIP)
  - Ilość „KlientN APP" nie ma zaznaczenia –licencjonowany jest tylko CTS względem CM.
- Ceny: CM – 14.4k Euro netto
  CTS – 8060 Euro netto
  CADP – 8060 Euro netto

  Opcja Flex – 450 Euro netto

Quiz z nagrodami!

Logujemy się, klikamy i nagrody wygrywamy!

https://forms.office.com/r/vsjLu8xmGG

CryptoPanel