UTOPIA FRAMEWORK - TOKENIZATION CONCEPTS

The Case for Tokenization

Abstract

Target/Victim modeling is as important as Adversary/Attacker modeling.

Automated Tokenization allows us to share categorical and quantitative data that can be used effectively for analysis and modeling.

Patrick Maroney

pmaroney@specere.org

Utopia Framework - Tokenization Concepts

OVERVIEW

Use Cases for Automated CTI Obfuscation, Tokenization, Redaction.

When sharing Attacker TTPs, especially in a real-time community of trust within a given sector/target group, quickly sharing detailed log events of the Attacker targeting patterns/timing between organizations targeted in a campaign are very valuable for a number of analyst processes.

(1) **Targeting Analysis**: Are they targeting organizations in succession, if so, in what order/timing? Is there any pattern to the individual target ordering? Did the targets attend or plan to attend the same conference, are they all members of a common professional organization, do they work in similar fields, roles, technologies?

(2) **Attribution**: Does the attack pattern match prior multi-organizational distribution list/ordering for prior Campaigns.

(3) **Attacker Tool Set Fingerprinting**:
   a. Tool "x" sends each email separately,
   b. Tool "y" sends each email in blocks of 128 addressees,
   c. Tool "z" sends one email every 10 seconds.

(4) **Early warning**: Alignment of large target lists to prior TTPs provide high certainty Identification of new campaigns by actor "x".

Since these detailed logs contain specific Employee, Infrastructure, Organizational target details, most organizations will need to redact, tokenize, obfuscate portions of the detailed logs before externally sharing.

MANUAL PROCESSING

Manual redaction/review processes can delay threat intelligence vital to the identification, characterization, and prioritization of emerging targeted attacks/threats.

- Manual Tokenization and Redaction processes take resources better spent on analytical and operational mitigation activities. They introduce potential leakage risks.

- Manual processing of log data does not scale well to sharing multi-columnar event data with 10,000s of rows.

REDACTION

- Target/Victim data is often redacted or removed completely from shared CTI packages to prevent leakage/exposure of organization attributional data.

- While redaction/deletion can be effective for sharing some key details, the loss of fidelity in our collective "eyes on" Victim/Target dynamics leaves us blind to key patterns in Attacker/Adversary TTPs.

Utopia Framework - Tokenization Concepts

CONCEPTS

Target/Victim modeling is as important as Adversary/Attacker modeling.

In fact, Victims/Targets may be the most critical half of the Cyber Battle Space.

Automated Tokenization allows us to share categorical and quantitative data that can be used effectively for analysis and modeling.

Target/Victim modeling is key to tactical objectives like reconnaissance detection/characterization, attack packaging/targeting analysis, attribution, etc. These model elements are also required to drive strategic processes like Predictive Analytics.

Providing a rich standardized taxonomy for Target/Victim enables key methods like redaction, tokenization, target generalization to aggregate Sector/Technology specific intelligence, etc. An effective cross-sector Business_Function taxonomy is a critical pre-requisite to non-attributional tokenization, conveyance, and modeling of Adversary Targeting TTPs.

(1) Targeted_Business_Function
(2) Targeted_Information
(3) Targeted_Asset_Class
(4) Targeted_System_Function

TOKENIZATION USE CASE

SCENARIO: TARGETED SPEAR PHISHING ATTACK

We will use an increasingly common scenario across sectors and companies targeted for coordinated exploitation by Determined Adversaries. We will apply relatively simply Tokenization methods in this paper to demonstrate core concepts. More complex methods for mapping Business_Function Taxonomies and Targeting Generalizations will be covered in a related paper.

Organization BigCo receives a series of malicious emails targeting it's employees. At the time nothing malicious is detected, or flagged as suspicious, so they are delivered to all addressees.

The attack is initially discovered after one of the targeted employees calls the Help Desk to report issues with Remote VPN Access. She claims she is following instructions she received via the email and going to the Link in the Email. The web site appears identical to the Self Service PIN/Password reset employees use frequently. She reports that she was asked to enter her account logon credentials, email address, answer Security Questions, and then instructed to enter the Next Token Code. She's then instructed to reset her Software Token PIN and VPN Password. After successfully completing the resets, she finds she can no longer logon to the Corporate Network using her VPN Credentials. She repeats the actions at the Web Site used to reset her PIN/Password, but still cannot logon. She eventually gives up and calls the Help Desk Hotline directly to report the problem (the spoofed email has the correct Help Desk Hotline information).

As they start to work with the US located employee and review VPN logs, they detect multiple authenticated active sessions under her ID originating from Germany and India.

Utopia Framework - Tokenization Concepts

The Help Desk escalates the issue to the CSIRT Team who quickly discover suspicious emails spoofing their Support Mail Distribution List. 1,000s of employees were targeted. A small sample is shown:

Date	Host	Sender	Recipient	Subject	Last State
2016-03-17 18:43 GMT	IRONPORT122 166.21.244.1	dil0712@megaorigamistar.info	bill.bates@bigco.com	Subj: Important - Reset Your VPN password	Message 282501216 to exchange_bridgehead@hq.bigco.com received remote SMTP response 'Queued'
2016-03-17 18:44 GMT	IRONPORT122 166.21.244.1	dil0712@megaorigamistar.info	bill.bates@hq.bigco.com	Subj: Important - Reset Your VPN password	Message 282501217 to exchange_bridgehead@hq.bigco.com received remote SMTP response 'Queued'
2016-03-17 18:45 GMT	IRONPORT122 166.21.244.1	dil0712@megaorigamistar.info	yolanda.lamda@bigco.com	Subj: Important - Reset Your VPN password	Message 282501218 to exchange_bridgehead@hq.bigco.com received remote SMTP response 'Queued'
2016-03-17 18:46 GMT	IRONPORT122 166.21.244.1	dil0712@megaorigamistar.info	charlie.baker@bigco.com	Subj: Important - Reset Your VPN password	Message 282501219 to exchange_bridgehead@hq.bigco.com received remote SMTP response 'Queued'
2016-03-17 18:47 GMT	IRONPORT122 166.21.244.1	dil0712@megaorigamistar.info	support@support.bigco.com	Subj: Important - Reset Your VPN password	Message 282501220 to exchange_bridgehead@hq.bigco.com received remote SMTP response 'Queued'
2016-03-17 18:48 GMT	IRONPORT122 166.21.244.1	dil0712@megaorigamistar.info	sales@hq.bigco.com	Subj: Important - Reset Your VPN password	Message 282501221 to exchange_bridgehead@hq.bigco.com received remote SMTP response 'Queued'
2016-03-17 18:49 GMT	IRONPORT122 166.21.244.1	dil0712@megaorigamistar.info	jsmith@dev.bigco.com	Subj: Important - Reset Your VPN password	Message 282501222 to exchange_bridgehead@hq.bigco.com received remote SMTP response 'Queued'
Gap in activity between 18:49 GMT 22:00 GMT					
2016-03-17 22:00 GMT	IRONPORT122 166.21.244.1	Help.Desk@pchelps.com	bill.bates@bigco.com	Subj: Important Account Notice	Message 282501433 to exchange_bridgehead@hq.bigco.com received remote SMTP response 'Queued'
2016-03-17 22:01 GMT	IRONPORT122 166.21.244.1	Help.Desk@pchelps.com	bill.bates@hq.bigco.com	Subj: Important Account Notice	Message 282501434 to exchange_bridgehead@hq.bigco.com received remote SMTP response 'Queued'
2016-03-17 22:02 GMT	IRONPORT122 166.21.244.1	Help.Desk@pchelps.com	yolanda.lamda@bigco.com	Subj: Important Account Notice	Message 282501435 to exchange_bridgehead@hq.bigco.com received remote SMTP response 'Queued'

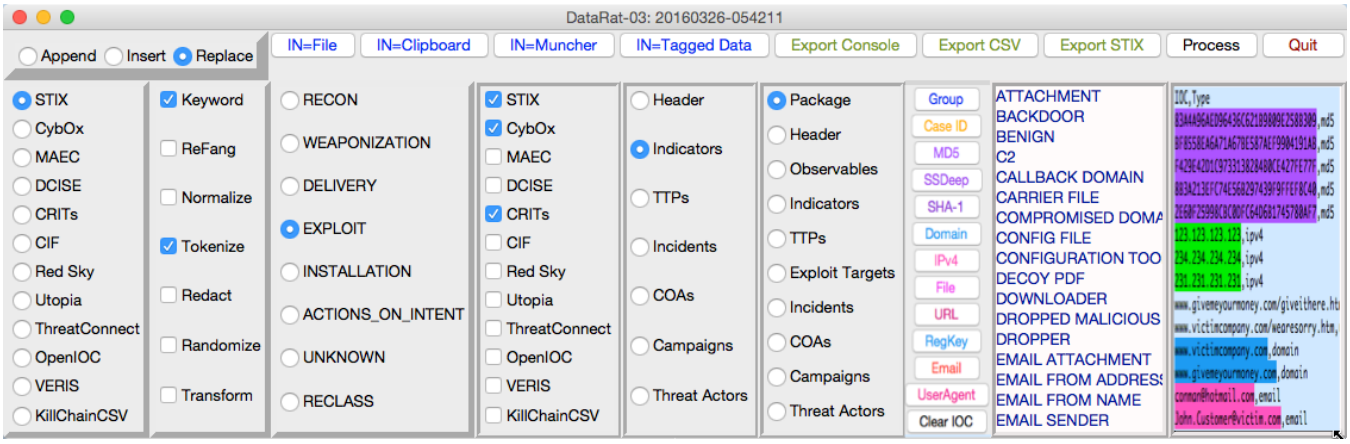
Preliminary investigation indicates that the attacker is evidently able to exploit an unknown 0Day vulnerability to re-generate Soft Tokens for existing user accounts without the requisite registered mobile device. All indications are the adversaries are commandeering employee VPN Accounts and successfully using them for access to internal sensitive systems/data networks.

The Authentication/VPN products are known to be widely deployed across BigCos Sector. At this point the only thing that is known is (1) a set of malicious emails were delivered to 1,000s of BigCos employees, (2) Unknown unauthorized actors are accessing BigCos "stuff".

Utopia Framework - Tokenization Concepts

Prior to CTI adoption and automatic Tokenization, all BigCOs resources would have focused on the mitigating the current operational exposures before taking time to manually review/edit/redact logs for external sharing.

In the Utopia Framework, the CSIRT Team immediately transmits 1,000s of Tokenized Mail Gateway log entries back to start of known malicious activity to present time to all members of it's ISAC.



Due to automated Tokenization and increased adoption of CTI STIX, the CSIRT team can also quickly send the initial investigation findings and raw log data to US-CERT, and to the NCI to distribute to it's members in other sectors.

Date	Host	Sender	Receipient	Subject	Last State
2016-03-17 18:43 GMT	Node-0003.Domain-0001 10.0.0.1	dil0712@megaorigamist ar.info	Invalid-address@Domain-001	Subj: Important - Reset Your VPN password	Message 282501216 to Node-0002.Domain-0005 received remote SMTP response 'Queued'
2016-03-17 18:44 GMT	Node-0003.Domain-0001 10.0.0.1	dil0712@megaorigamist ar.info	Name-0002@Domain-0005	Subj: Important - Reset Your VPN password	Message 282501217 to Node-0002.Domain-0005 received remote SMTP response 'Queued'
2016-03-17 18:45 GMT	Node-0003.Domain-0001 10.0.0.1	dil0712@megaorigamist ar.info	Name-0001@Domain-0001	Subj: Important - Reset Your VPN password	Message 282501218 to Node-0002.Domain-0005 received remote SMTP response 'Queued'
2016-03-17 18:46 GMT	Node-0003.Domain-0001 10.0.0.1	dil0712@megaorigamist ar.info	Name-0003@Domain-0001	Subj: Important - Reset Your VPN password	Message 282501219 to Node-0002.Domain-0005 received remote SMTP response 'Queued'
2016-03-17 18:47 GMT	Node-0003.Domain-0001 10.0.0.1	dil0712@megaorigamist ar.info	Name-0004@Domain-0004	Subj: Important - Reset Your VPN password	Message 282501220 to Node-0002.Domain-0005 received remote SMTP response 'Queued'
2016-03-17 18:48 GMT	Node-0003.Domain-0001 10.0.0.1	dil0712@megaorigamist ar.info	Name-1248@Domain-0005	Subj: Important - Reset Your VPN password	Message 282501221 to Node-0002.Domain-0005 received remote SMTP response 'Queued'
2016-03-17 18:49 GMT	Node-0003.Domain-0001 10.0.0.1	dil0712@megaorigamist ar.info	Name-1250@Domain-0003	Subj: Important - Reset Your VPN password	Message 282501222 to Node-0002.Domain-0005 received remote SMTP response 'Queued'
2016-03-17 22:00 GMT	Node-0003.Domain-0001 10.0.0.1	Help.Desk@pchelps.co m	Invalid-address@Domain-001	Subj: Important Account Notice	Message 282501433 to Node-0002.Domain-0005 received remote SMTP response 'Queued'
2016-03-17 22:01 GMT	Node-0003.Domain-0001 10.0.0.1	Help.Desk@pchelps.co m	Name-0002@Domain-0005	Subj: Important Account Notice	Message 282501434 to Node-0002.Domain-0005 received remote SMTP response 'Queued'
2016-03-17 22:02 GMT	Node-0003.Domain-0001 10.0.0.1	Help.Desk@pchelps.co m	Name-0001@Domain-0001	Subj: Important Account Notice	Message 282501435 to Node-0002.Domain-0005 received remote SMTP response 'Queued'

It's a "One-Clicker" so they also forward a copy of the preliminary STIX Report to the Authentication and VPN Vendors.

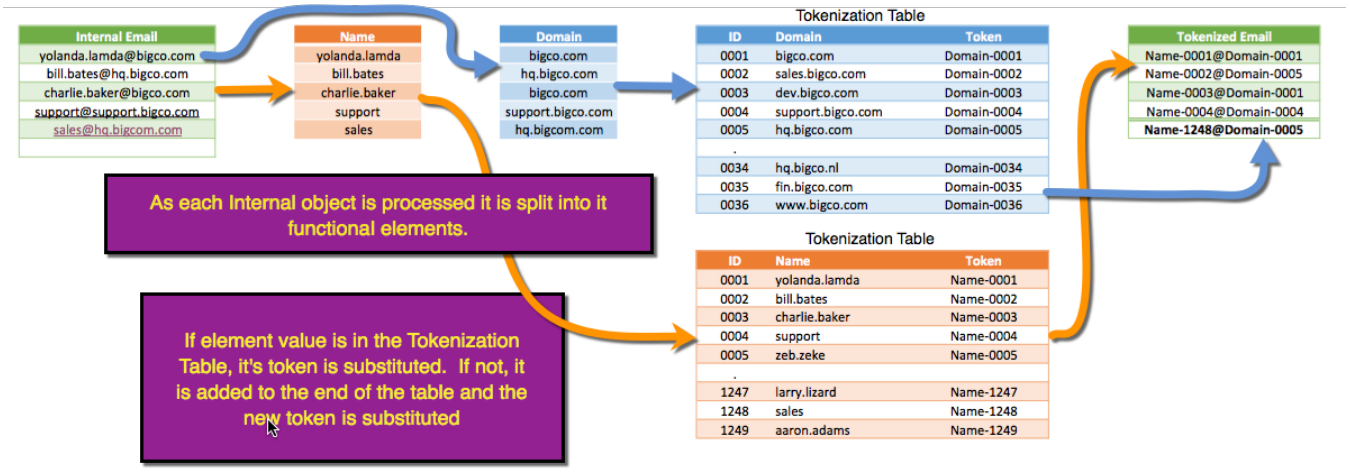
Utopia Framework - Tokenization Concepts

Cross correlation of tokenized email logs from multiple Analysts across targeted organizations reveal multiple waves of attacks, and samples of multiple variants of the Attack Package and malware exploit.

Matching to previous cross-sector targeting lists makes high certainty attribution to Nation State Actor "X", also known for aggressive and successful exploitation of VPN Technologies. This actor was also suspected of directly targeting company Help Desk Support Distribution Lists, but no evidence of actual malicious activity had ever been isolated.

Increasing the look-back period for cross correlation of 10,000s of tokenized mail log across the targeted companies quickly reveals that two weeks prior to the main attack, a series of similarly themed emails to most of the external Help Desk DLs of targeted companies. Emails all originated from Employees claiming to be on travel, having VPN issues, and requesting the latest copy of that company's VPN User Guides. By correlating logs and targeting it was concluded that these earlier emails from Adversary were battle space preparation to capture up to date VPN Self Service Screenshots explaining how they were able to effectively spoof VPN Self Service Web Portals.

ORGANIZATIONAL EMAIL ADDRESS AND DOMAINS



- The Email Name and Domain are split.
- If Email Name is not in Email Name Token Table (1) insert it and (2) generate Email Name Token:

  "Name-" & Email Name Token ID

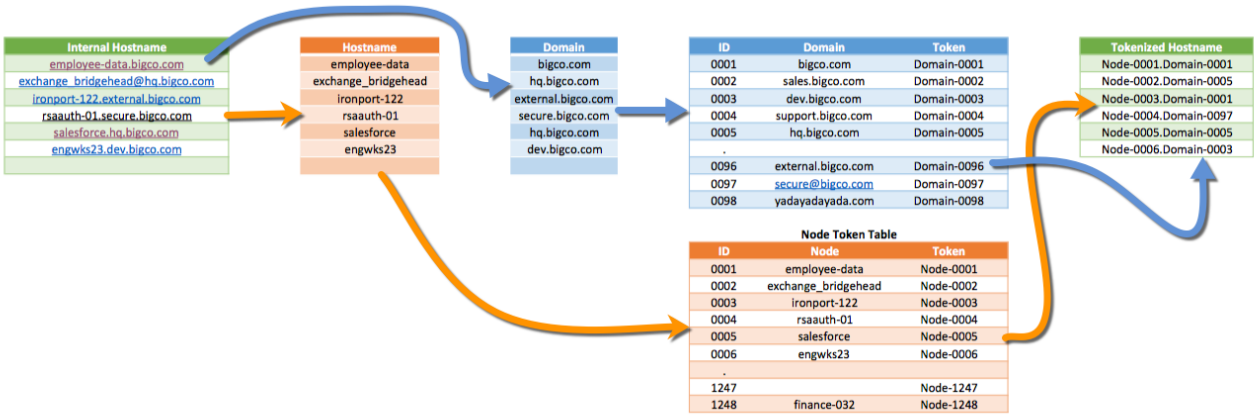
- If Domain Name is not in Domain Name Token Table (1) insert it and (2) generate Domain Name Token:

  "Domain-" & Domain Name Token ID

Utopia Framework - Tokenization Concepts

NETWORK OBJECTS

NETWORK NODES - NAMES & DOMAINS



- The Host Name and Domain are split.

- If Host Name is not in Host Name Token Table (1) insert it and (2) generate Host Name Token:

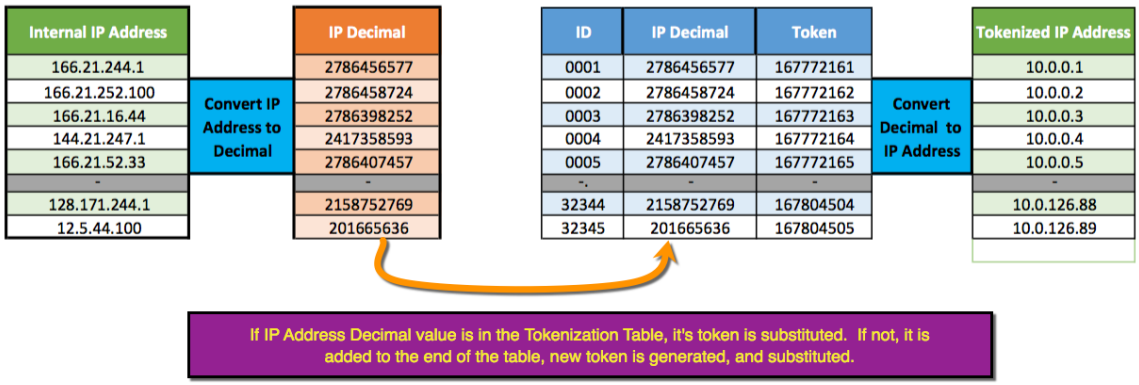
  "Host-" & Host Name Token ID

- If Domain Name is not in Domain Name Token Table (1) insert it and (2) generate Domain Name Token:

  "Domain-" & Domain Name Token ID

Utopia Framework - Tokenization Concepts

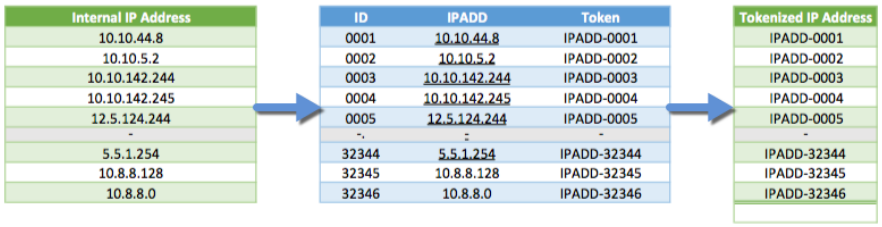
NETWORK NODES – IP ADDRESSES



- The IP Address is converted to decimal.
- If IP Address is not in IP Address Token Table, (1) insert it and (2) generate IP Address Token

DecimalToIP((IP Address Token ID + (IPDecimal(10.0.0.1))

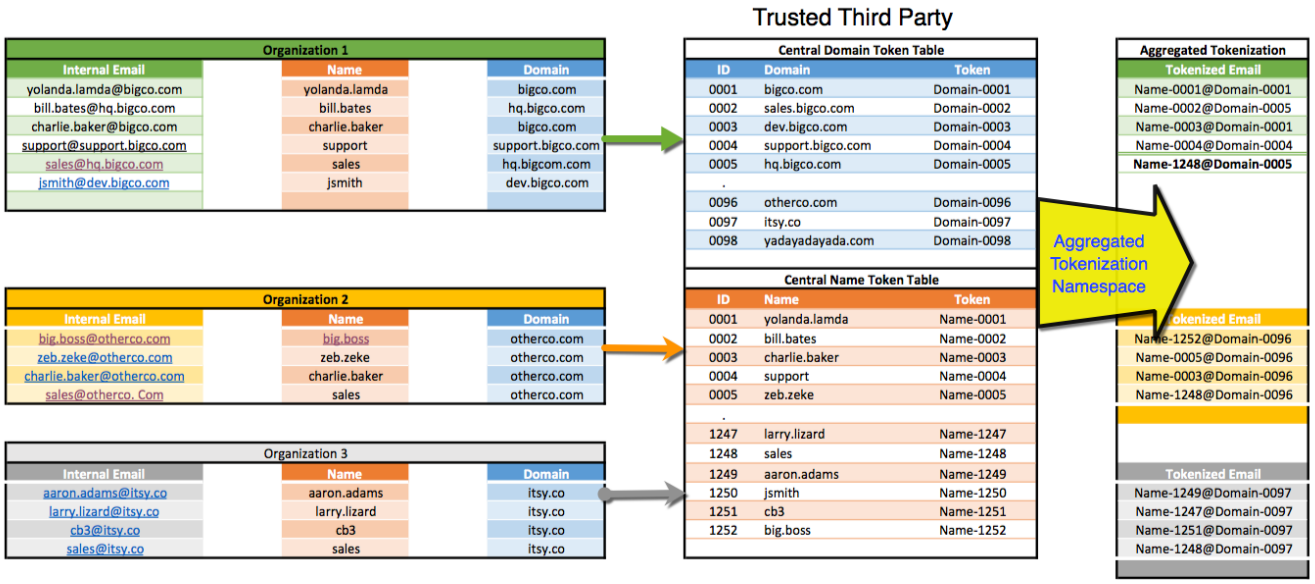
An Alternate IP Address Tokenization Method



Utopia Framework - Tokenization Concepts

TOKEN AGGREGATION/DISEMINATION

Another Tokenization Use Case is when multiple organizations individually report incidents to a trusted third party/central agency. The trusted agent generates central token mapping tables for all participants and provides these mappings in the CTI it shares with contributing organizations and other stakeholders. The central agency can provide a consistent set of tokenization mappings to consistently describe key targeting data for all the events in a global context and over extended periods of time.



OBFUSCATION - RECOVERABLE

Another mapping/transformation method uses a random or algorithmically generated mapping table. In the IP Address Obfuscation Tokenization Table example shown here the entire IPV4 Space is represented and specifically mapped, tuple for tuple, address to address.

The table here was generated algorithmically, Which provides for selectively sharing the Transformation Mappings required to recover the real IP Address Value.

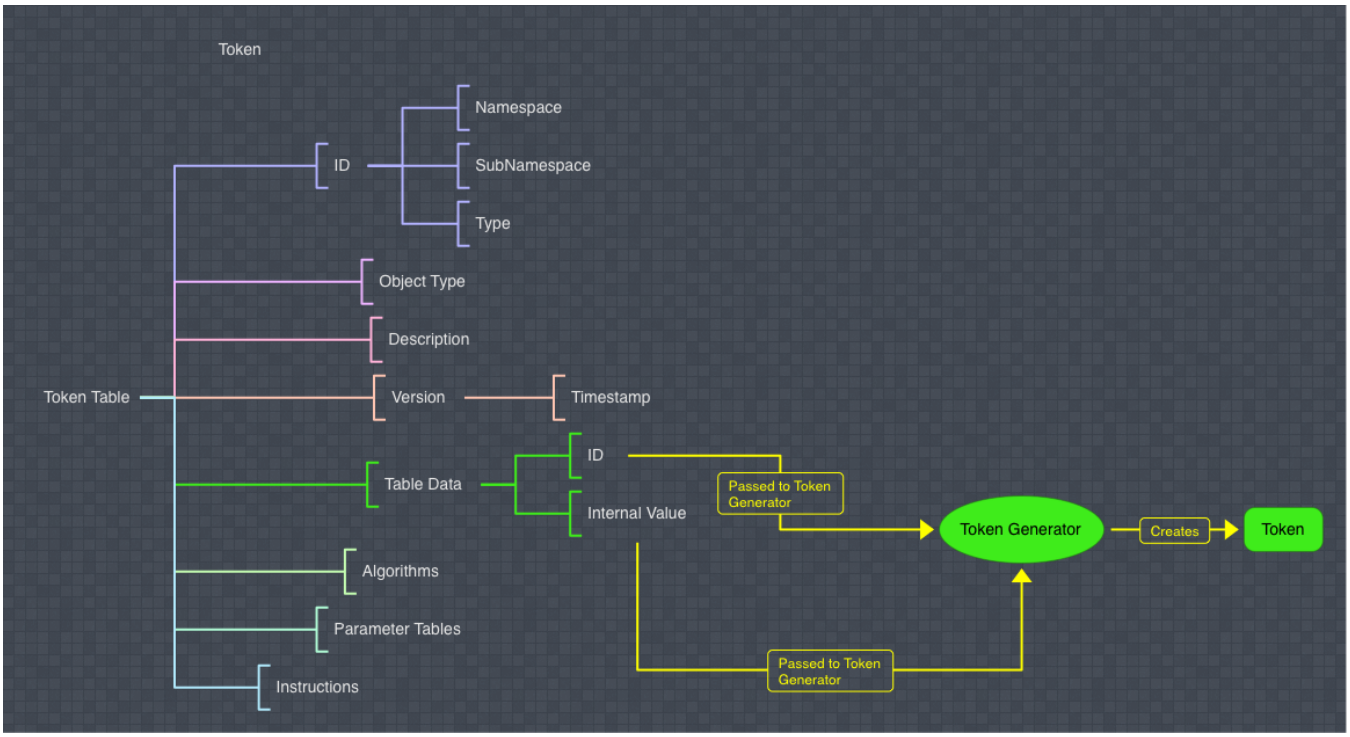
This technique can also be applied when concerned with externally exposing highly sensitive Adversary/Attacker Addresses. These values can Safely passed "hiding in plain sight".

Utopia Framework - Tokenization Concepts

TOKEN STRUCTURE

The notional Token Table Structure is depicted below. A common Table Data Type is generally required for conveying Textual Row/Column data (with Headers), and specifically required for inter-exchange of populated Token Tables, Parameter Tables contained within Token Tables, and Tokenized Log/Event Data.

The Token Table ID is generated using *uuidv5 Hash of the Source Namespace/sub-namespace and Immutable Object Properties [Token Table Type, Object Type].*



A modular scriptable language is foreseen providing complex, multi-state, multi-stage Tokenization Algorithms and Token Generators. Existing frameworks (e.g., OpenRefine, Python) are currently being applied for these methods/functions.

Parameter Tables Support Multi-Stage processes.

For example, the IP Address Tokenization Table that maps successive IP Addresses to sequential 10.0.0.0 space includes two conversion processes (1) **Convert IP to Decimal** and (2) **Convert Decimal to IP**. Both Internal IP and Token
Decimal representations must be stored in Parameter Table Space.

Internal IP Address
166.21.244.1
166.21.252.100
166.21.16.44
144.21.247.1
166.21.52.33
-
128.171.244.1
12.5.44.100

Convert IP Address to Decimal

IP Decimal
2786456577
2786458724
2786398252
2417358593
2786407457
-
2158752769
201665636

ID	IP Decimal	Token
0001	2786456577	167772161
0002	2786458724	167772162
0003	2786398252	167772163
0004	2417358593	167772164
0005	2786407457	167772165
-	-	-
32344	2158752769	167804504
32345	201665636	167804505

Convert Decimal to IP Address

Tokenized IP Address
10.0.0.1
10.0.0.2
10.0.0.3
10.0.0.4
10.0.0.5
-
10.0.126.88
10.0.126.89