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

THE CASE FOR AUTOMATED CTI TOKENIZATION.

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" User-Agent = "RSX-Mail 1.5.0.7 (DEC PDP-11/34)".

(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 also contain attributional details on Employees, Organization, and Infrastructure, organizations need to redact these portions of these logs before externally sharing. These redaction/review processes:

- Strip away very useful details on the Attack, Attacker/Intermediaries/Targets, and TTPs.
- Delay the dissemination of threat intelligence vital to the identification, characterization, and prioritization of emerging targeted attacks/threats.
- Consume resources better spent on analytical and operational mitigation activities.
- Increase risk of unintentional exposures of attributional data.
- Do not scale for sharing large amounts of event log data (i.e., 1,000s of rows of multi-columnar data).

Utopia Framework - Tokenization Concepts

CONCEPTS

Targets/Victims are as important as Adversaries/Attackers. In fact, Target/Victims may be one of the most critical elements of effective Cyber Battle Space modeling.

The automated tokenization and sharing of today's highly redacted Target/Victim data would provide key missing categorical and quantitative variables required for complex dynamic system modeling and analytics.

For example, Target/Victim modeling is key to the detection and characterization of the following CTI categories (both at a given time, and how they change over time)

(1) Reconnaissance.
(2) Attacker objectives.
(3) Attack package construction/delivery methods.
(4) Targeting list construction/targeting methods.
(5) Attack Patterns.
(6) Metrics: "How do my mitigation Course of Actions compare to similar companies in my sector?"
(7) Predictive Analytics: *"How will Adversary "x" respond to event "y", and When?*')

In addition to the system and network level tokenization methods outlined in this Use Case, the development of broader Sector/Technology specific taxonomies would enable the "smart-redaction"/tokenization necessary to aggregate Sector/Technology specific intelligence.

(1) Business Function
(2) Information
(3) Asset Class
(4) System Function

The development of effective cross-sector taxonomies would in turn enable similar tokenization, conveyance, and target generalization capabilities necessary to identify Advanced Adversaries operating in a Global context and the evolution of their Targeting TTPs over time and across sectors.

Utopia Framework - Tokenization Concepts

TOKENIZATION USE CASE

We will use a common Determined Adversary Targeted Spear Phishing Attack scenario where multiple companies are targeted in multiple successive campaigns. We will apply relatively simple system and network level Tokenization methods to demonstrate core concepts.

SCENARIO: TARGETED SPEAR PHISHING ATTACK

Organization BigCo receives a series of emails as part of a broader campaign targeting 1,000s of employees at dozens of companies in BigCo's Sector. Nothing suspicious/malicious is detected by BigCo's incoming Email scanners at the time, so all messages are delivered to the employees.

Symptoms of the attack are first discovered when one of the targeted employees calls the Help Desk to report issues with Remote VPN Access. She reports that she is following instructions in an Email received from BigCO's Help Desk to reset her VPN Password via the web link in the message.

She reports that she was taken to the BigCo Self Service Password Reset Web Site that remote employees use frequently. She logged in successfully with her credentials and followed the instructions to reset her her VPN Password (the Software Token PIN). After completing the requested actions, she was taken to the "usual" Corporate Network Logon screen, where she entered her Windows Username/Password, but is unable to connect to Outlook Server or access any services.

She repeats the instructions at the Web Site in the email to reset her PIN/Password, but still cannot logon to the Corporate Network. She eventually gives up and calls the Help Desk Hotline directly to report the problem (along with what appears to be the 'correct' link to the VPN Self Service portal, the spoofed email also has the correct Help Desk Hotline phone information).

As the Help Desk starts to work with the employee, they verify her US location/IP address and begin a review of VPN logs. They quickly detect multiple active VPN sessions under her ID originating from Germany and India.

The Help Desk immediately escalates the issue to BigCo's CSIRT Team

Utopia Framework - Tokenization Concepts

The CSIRT Team quickly isolate the initial set of related emails and determine they are from an external address spoofing their VPN Support Team email address.

They determine 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 able to re-generate Soft Tokens for existing user accounts without the requisite registered mobile device. All indications are the adversaries are actively commandeering employee VPN Accounts and using them for access to internal sensitive systems/data networks.

At this point what is known for sure is:

(1) Emails were delivered to 1,000s of BigCo employees with instructions to immediately reset VPN passwords via an embedded link to a site spoofing BigCo's VPN Self Service Portal.

(2) Employees are clicking this link and providing credentials.

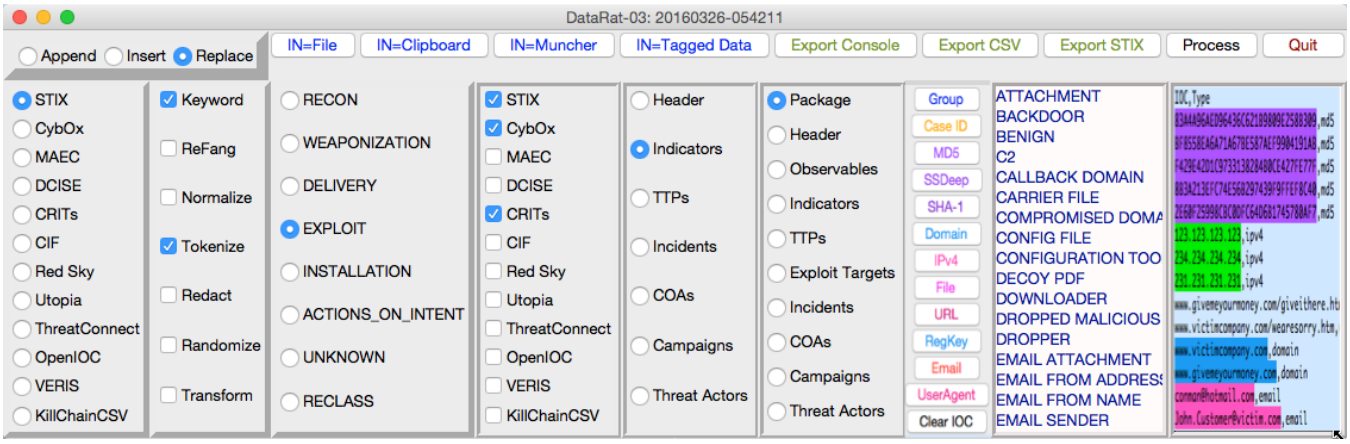
(3) Using credentials harvested from this spoofed site, unidentified external actors are able to
(3.1) Re-provision Soft Tokens without requisite Mobile Device Hardware Keys (presumed 0Day)
(3.2) Re-provision existing VPN Access User Accounts
(3.2) Remotely access BigCo internal systems and data

(4) The Authentication/VPN products in question are widely deployed at companies across BigCo's Sector.

Utopia Framework - Tokenization Concepts

Prior to the adoption of automated Interexchange and Tokenization using OASIS CTI STIX, CybOX, and TAXII Standards, all of BigCo's resources would have focused on mitigating the current operational exposures before taking time to manually review/edit/redact logs for external sharing.

In the OASIS CTI empowered Utopia Framework, the CSIRT Team immediately selects and transmits the initial investigation findings along with the Tokenized Mail Gateway log entries for the malicious activity isolated to date 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 with tokenized 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.

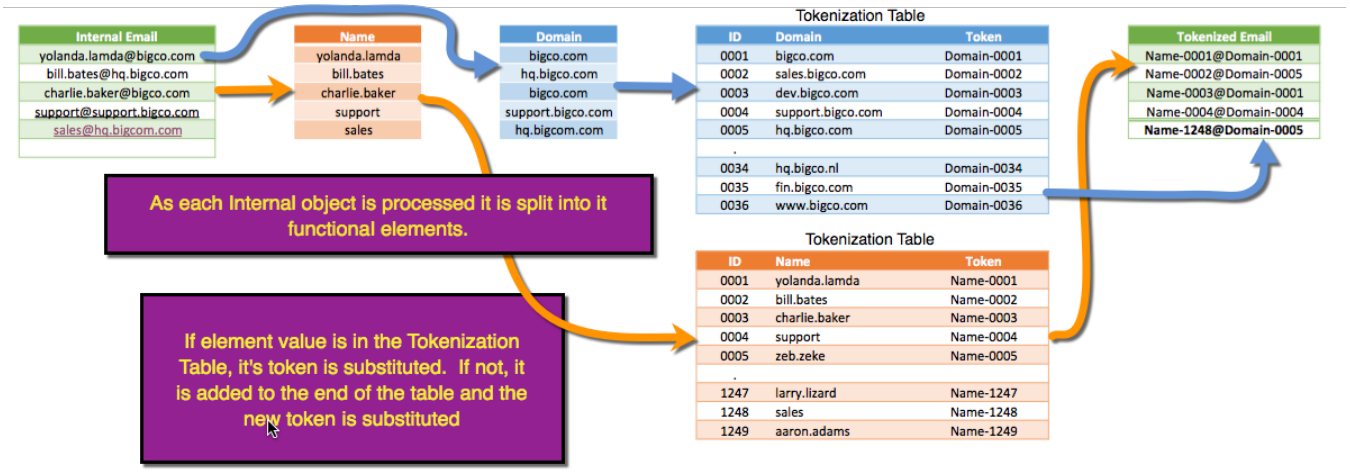
Utopia Framework - Tokenization Concepts

The cross correlation and analysis of tokenized email logs shared between Analysts of the targeted organizations reveal multiple waves of attacks. Analysis and pivoting on this data provides samples of multiple variants of Attack Packages and samples of the 0Day exploit used against VPN Appliances and Soft Token Provisioning functions of the Authentication Servers. Early engagement with the VPN and Authentication vendors at the initial discovery of the attack aids in identification of the exploits, short term mitigations, and delivery of a fully QA'd Patch.

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 and cross correlation of 100,000s of tokenized mail logs from the targeted companies reveals that two weeks prior to the main attack, a series of similarly themed emails to 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 "one-off" emails from Adversary were battle space preparation to capture up to date VPN Self Service Screenshots potentially 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

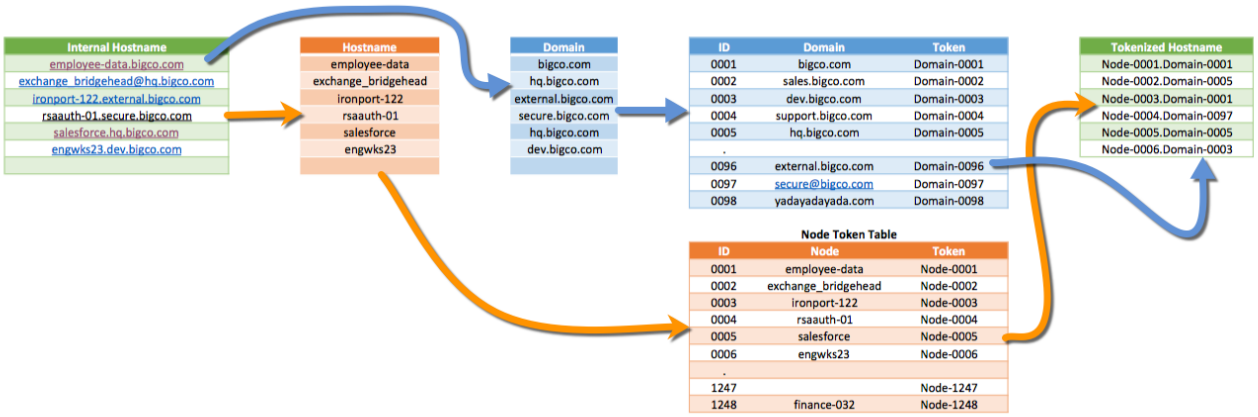
- If Domain Name is not in Domain Name Token Table (1) insert it and (2) generate Domain Name Token:

  "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:

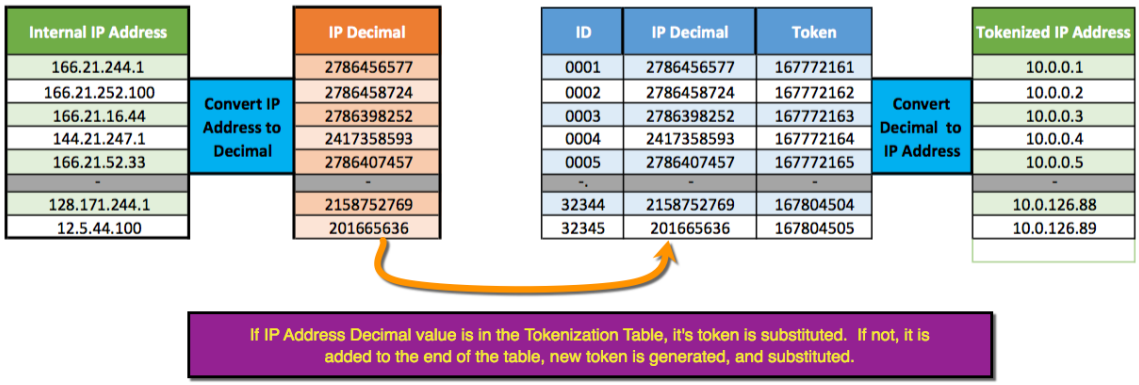
  "Host-" & Host Name Token ID

- 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

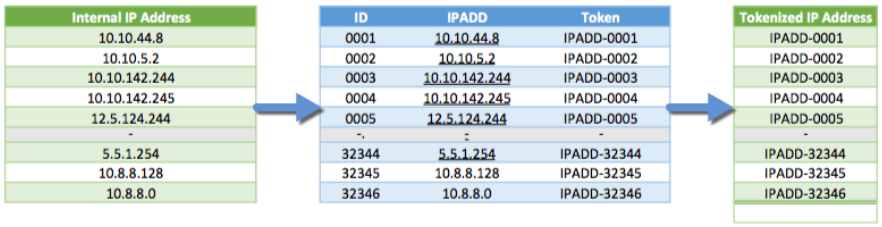
NETWORK NODES – IP ADDRESSES



- 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))

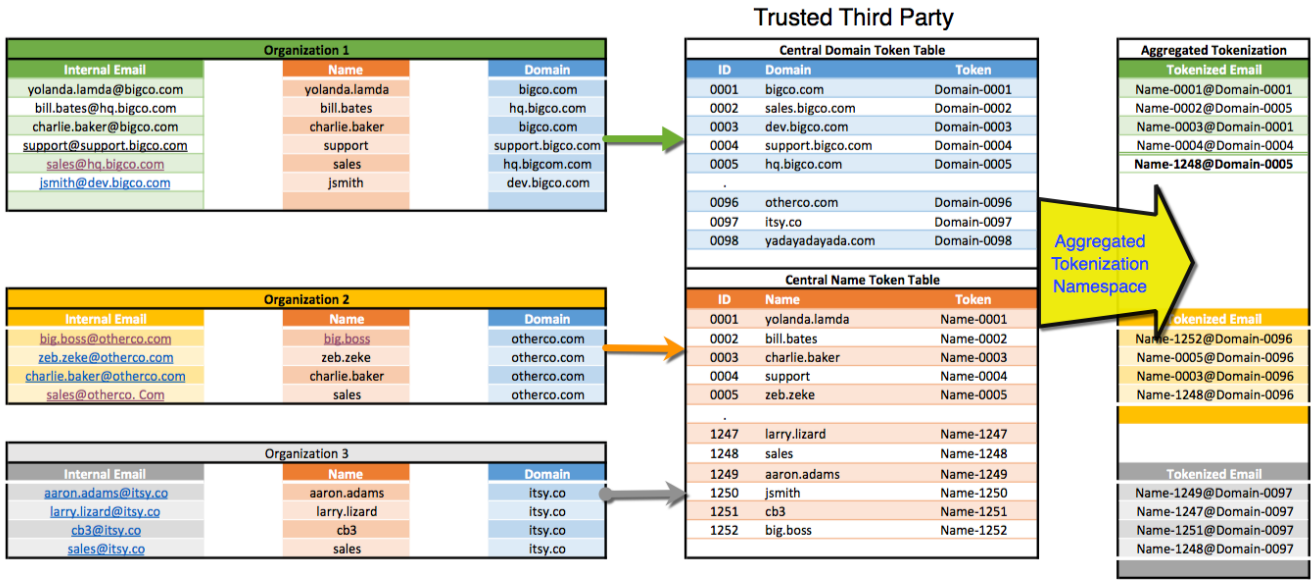
An Alternate IP Address Tokenization Method



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.

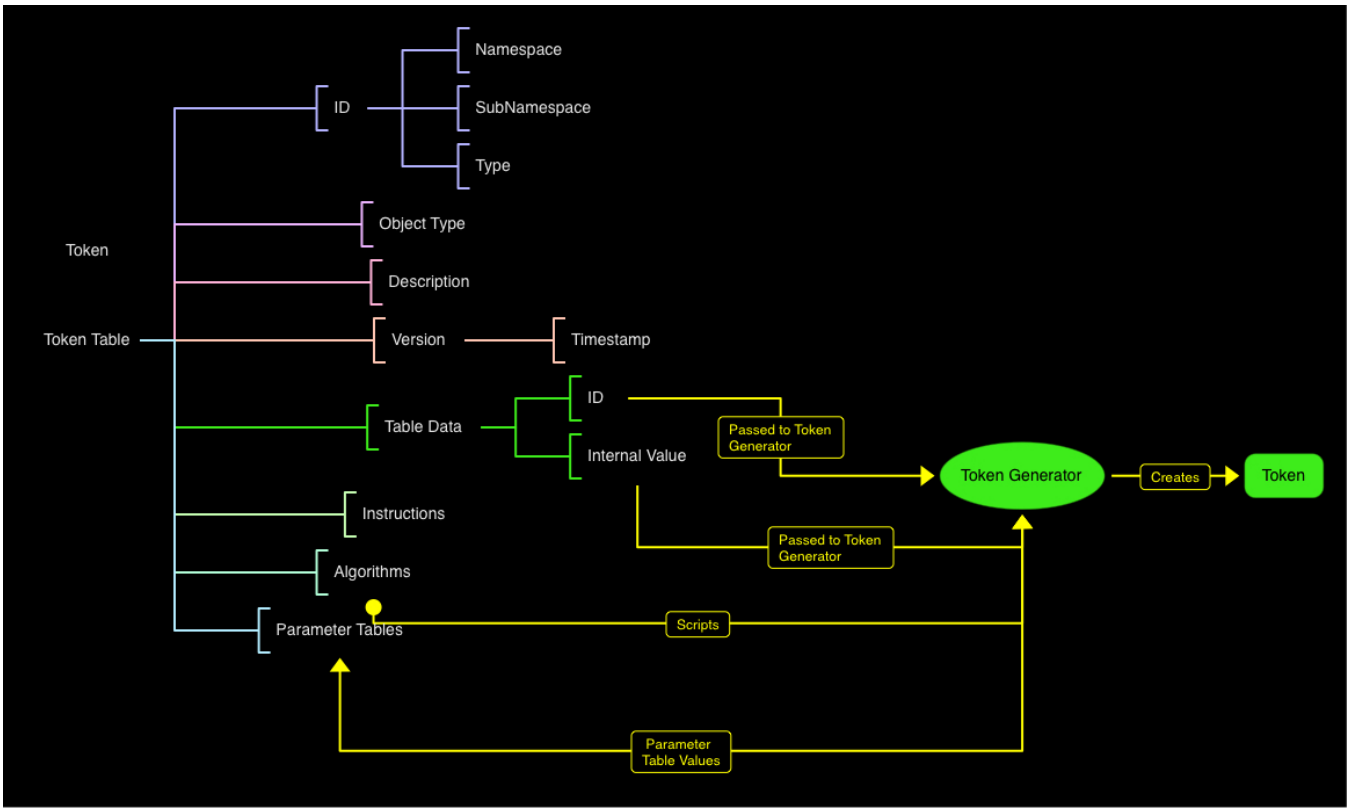
This technique can also be applied when concerned with externally exposing highly sensitive Adversary/Attacker Addresses. These values can be safely passed "hiding in plain sight".

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/Pandas) 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