virtru

Zero Trust Analytics for Federal Customers

Executive Summary

As data has become an increasingly valuable enterprise asset, two critical missions have become more of a priority in an organization: data scientists and analysts continually call for increased visibility and access to data from across the organization, as well as the ability to combine this with data from external sources; while cybersecurity personnel continue to mitigate increasing levels of risk in the modern IT landscape. When organizations need to move to or maintain a Zero Trust security posture, while still enabling advanced analytics on sensitive data, Virtru's Cleanroom offers a Kubernetes-based platform to support both teams' needs.

Problem Statement

Over the last decade, organizations have increasingly considered data to be a significant enterprise asset. This has led to two trends, which are often seen in direct opposition to one another. First, significant benefit can be derived from detailed and often advanced analysis of data from across the organization or from mission partners. Data scientists and other analysts can create predictive models, answer questions relevant to business decisions, and guide the direction of the enterprise in meaningful ways by using data.

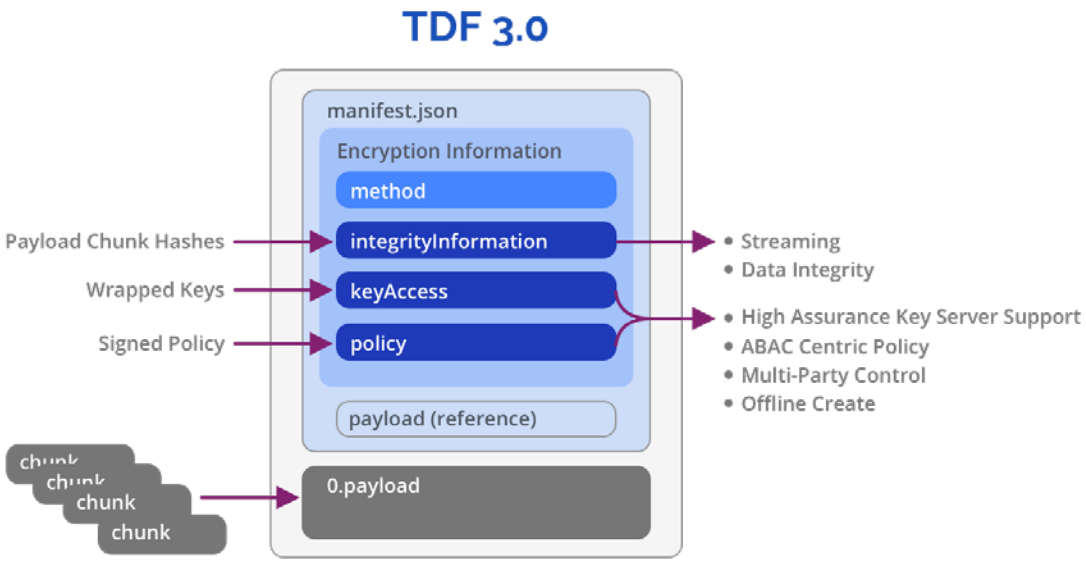
However, cybersecurity professionals express concern over the growing risk external threats pose to this enterprise asset. The level of openness desired by data scientists and analysts may be seen by the security team as an unacceptable level of risk. The emergence of Zero Trust models of security, and particularly the recent Executive Order 14028, Improving the Nation's Cybersecurity[1], mandating their use in Federal organizations, further emphasizes the importance of these two missions working together.

With Virtru's Cleanroom, an analytic environment that is trusted and secure becomes available to data scientists and other analysts, allowing them to use the tools they need while giving data owners complete control over both the source data and the *derived analytic products produced*. This complete control over access to both source data and derived analytical products even transcends the boundaries of the organization, giving complete control of data anywhere it is disseminated.

What Is Zero Trust?

At its core, Zero Trust is a shift in focus for IT security away from protecting *only* or *primarily* the network perimeter to protecting every element and piece of data within an IT system. Until now, once someone has gained access to the network, they are presumed to belong there and additional security hurdles are generally lower. When implementing Zero Trust, *every* component of the IT system is protected to the fullest extent: the network perimeter, each server, database, or other endpoint, and even *each individual piece of data*. In other words, should a nefarious actor gain access to a network or a database, they would only be able to cause minimal harm because *the data itself is individually protected*. Think of it as shrinking the security perimeter down until it surrounds each individual piece of data and each system endpoint.

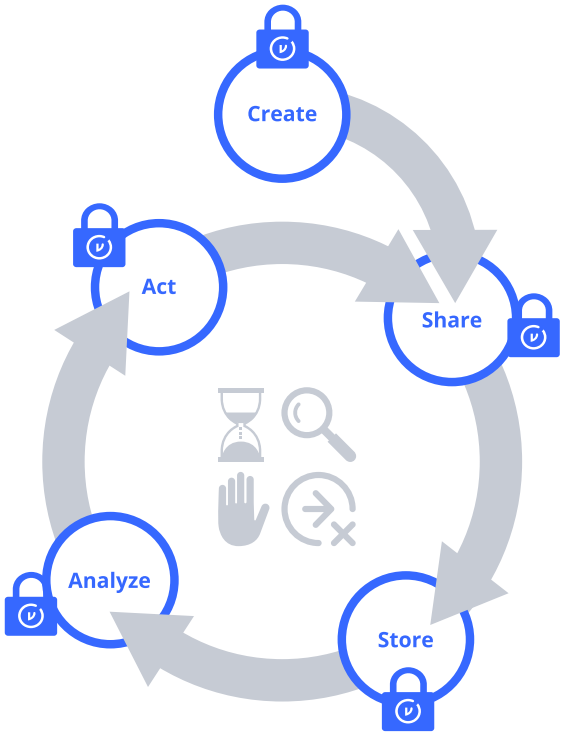
1 https://www.whitehouse.gov/briefing-room/presidential-actions/2021/05/12/executive-order-on-improving-the-nations-cybersecurity/



Virtru's Trusted Data Format (TDF), based on the ODNI-approved open standard, provides a protective wrapper that cryptographically binds protections to the data at the attribute level. TDF leverages Attribute-Based Access Control (ABAC) to enable customized and granular data access privileges that fulfill the least privileged access approach of a Zero Trust strategy. TDF is agentless and allows file locking, content expiration, and access revocation for both structured and unstructured data of any size, **regardless of the data's location — even after the data object has left the organization entirely**. Organizations and data owners can tag, encrypt, revoke, expire, and audit access to data, even after content has been accessed or after it has left the organization's systems.

How Can Analytics Work In A Zero Trust Environment?

When implementing a Zero Trust security paradigm, it isn't enough to simply consider query-response mechanics and how data will be viewed in such transactions. Data has truly become an *organizational asset*, one which has a clear business use. Throughout its entire lifecycle, from creation through analysis, dissemination, and re-analysis, how can data owners maintain protection of *their* data *wherever* it happens to be at the moment? And how can data owners determine which analyses are allowed on their data, and which are not? All of these considerations must factor into the security design.



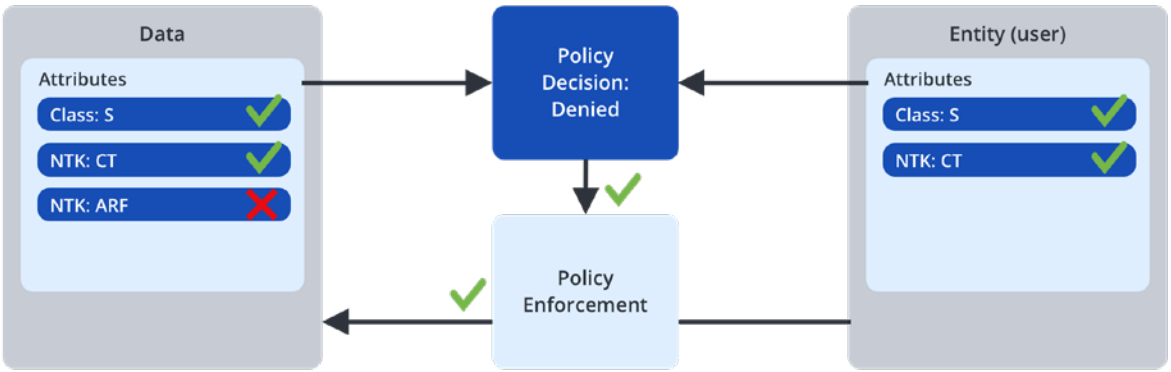
Data Creation

The workflow begins with categorizing and protecting data inputs. Data providers use client-side (local) applications and cryptographic keys to prepare data for input by wrapping it in the trusted data format (TDF). The TDF format contains encrypted data that is cryptographically bound to tags and associated access and handling policies. The TDF software development kit (SDK) is available to streamline client side data preparation. The client then sends the encrypted, tagged data to the trusted data platform.

Data Sharing – ABAC Attribute Management

ABAC is a logical access control methodology where authorization to perform a set of operations is determined by evaluating attributes associated with the subject, object, requested operations, and, in some cases, environment conditions against the policies, rules, or relationships that describe the allowable operations for a given set of attributes.

Before data can be accessed or used by another user, system, or analytic, data owners must approve that use of data. This can be done by name, by analytic type, or other attribute (e.g., role) describing the purpose or function of the analytic. Data owners assign entitlement attributes to their data, determining which systems and users have access under which conditions, as well as what analytic actions are permissible and which are not.



Storing Data – Applying Persistent Access Control

After protected data is released by the client, a cryptographically enforced Attribute-Based Access Control (ABAC) model is used to provide data owners with persistent controls to manage user and analytic access to their data and to its derived outputs. In the ABAC model, access decisions are made based on data attributes, policy logic, and user attributes. Users, systems, and analytics are assigned entitlements by data tag owners before they can gain access to a dataset. The TDF ensures that access to datasets is cryptographically enforced by encrypting all data and enforcing access policies at the key servers. To gain access to a decryption key, users must have strong identities, which have been assigned all applicable entitlements by the entitlement owners.

Secure Analytics

After data is encrypted and co-located, specific users or analytics may be approved to access the data by data owners. At the time of approval, analysts or analytics must define how the output or derived data will be tagged. In some cases, all of the tags of all the input data may be required to persist on the output. In other cases, data may be sufficiently summarized or obfuscated so that only some of the data tags are required. To ensure integrity of analytic logic, registered analytics are hashed prior to approval, and cannot be modified without being re-approved. In order for software to be granted access, it must be able to prove its identity and that its configuration state is secure.

Analytic Containers

After an analytic has been assigned a strong identity and has been approved to run over one or more datasets, it also needs a secure location to run. Analytic containers are secure machine images where analytic owners can run containerized software. The image provides cryptographic integrity measurements of the software image, as well as tamper-resistant configurations that are required by policy. All of these parameters are digitally signed and part of the identity used to request data access. Like analytics, Trusted Data Platform container identities can be required to be issued attributes by attribute authorities before they can be used to process data.

Tamper protections ensure that clear text is never allowed outside of the secure analytic container unless authorized, and that container configurations cannot be modified without forcing re-authentication via a trusted measurement and attestation process. Data owners approve what analytics can be run across their data, and they maintain full audit control.

Key Server Access Enforcement and Audit

In order for an analytic to access any encrypted data for processing, authorization must first be granted to the data encryption keys stored in the key servers. Data owners can independently control and audit the use of their data by managing the data tags and associated policies used by the key servers, even after the data is mixed with other data in a storage bucket, such as Amazon S3.

In order for analytics to request keys, each process is given a strong digital identity that can be used to connect to the data owner's key server. Data owners can even implement and fully control a key server to be used for their data. Data owners may choose to use on-premise, cloud, or third-party-hosted key servers.

Action From Data – Trusted Platform Outputs

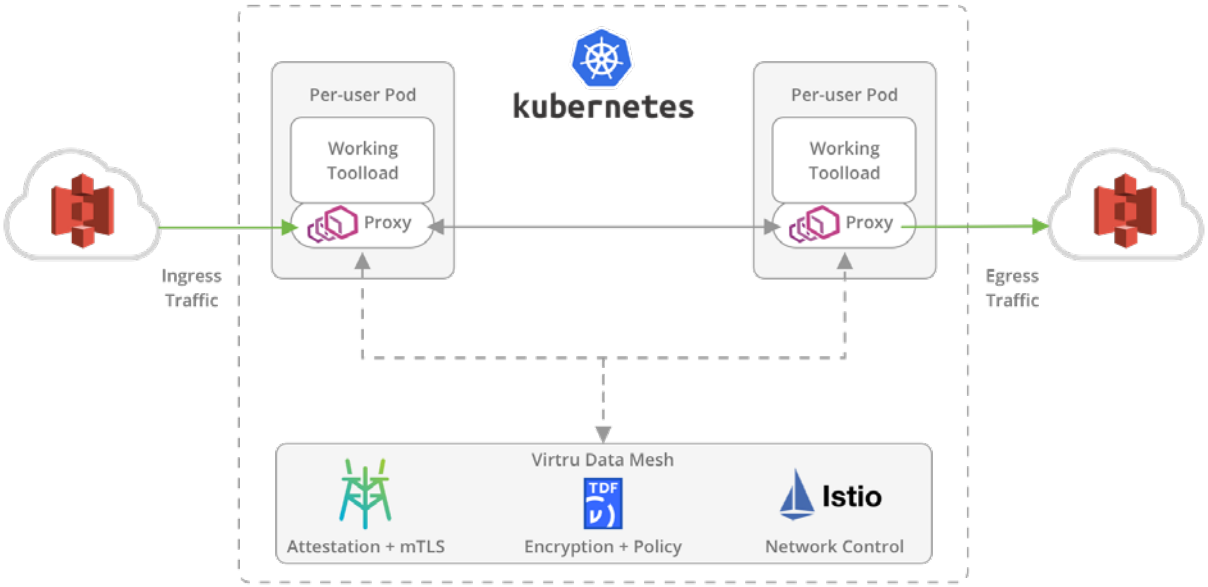
The Trusted Data Platform enforces encryption and protection of analytic outputs or derived data based on either manually approved output tags, or dynamically based on a combination of input data and analytic attributes. In particular, via data attribute drop/add conditions, policy language can support automated, dynamic output attribute tags. Analytic output will be encrypted based on those approved output tags, and only be accessible to entities that have been granted the approval to access data with all of the applied tags.

When an originator tag is required on output, the originator has the ability to audit all subsequent access to the derived data, change the access policy associated with their data tag, or revoke access.

The Value of Virtru Cleanroom

This Zero Trust analytic environment is implemented by Virtru's Cleanroom. This is a Kubernetes-based environment in which data engineers, data scientists, and other authorized analysts can securely operate on sensitive data and produce outputs which continue to be controlled by data owners who provided the input data. Cleanroom offers a number of benefits in its analytic environment:

- All derived data has a fully documented history of all steps involved in its creation and is traceable at a row level to the input data used in its creation.
- All data access and analytic actions produce a secure, verifiable audit trail for each data owner.
- All containers involved in the analytic process are cryptographically isolated, allowing individual data owners complete control.
- A flexible Kubernetes-based environment supports integration with unmodified (containerized) applications.
- All input and output to and from secure environments is only via supported SDKs.
- A Zero Trust data lake (usually implemented in S3, although other architectures are possible) allows for secure colocation of sensitive data.
- Tamper-resistant and provable identity of each container is available via attestation (e.g., via SPIFFE, SPIRE, TEE, etc.).



Next Steps

The heightened security posture required in today's cyber environment often leaves organizations wondering how they can fulfill both their security and analytic missions. With Virtru Cleanroom, both security and analytics can be enhanced.

As a critical part of the technical and cultural change required for Zero Trust, Virtru has been an important enabler for a data-centric vision and transformations to Zero Trust models. Virtru products have a proven history of protecting and sharing sensitive data across business and mission workflows for the U.S. intelligence community, the Department of Defense, other elements of the U.S. Federal government, and thousands of commercial customers. Virtru's Trusted Data Platform contains the necessary elements for implementing a modern Zero Trust strategy, and Virtru's implementation teams and professional services partners have the experience and expertise to guide Federal customers as they navigate this new and emerging security landscape.

Please contact federal@virtru.com and let our expertise support your mission.

Trusted by Federal Agencies, State and Local Governements.

At Virtru, we empower organizations to unlock the power of data while maintaining control, everywhere it's stored and shared. Our portfolio of encryption solutions and tools—built on our open data protection platform—governs data throughout its lifecycle. More than 6,000 customers trust Virtru for data security and privacy protection.