Multi-Level Security (MLS) Data Gateway for Agile and Secure Enterprise Data Sharing and Integration

Eltefaat Shokri, Josh Train, Ryan Waer,
Ann Chervenak, Scott Bergonzi,
Theodore Faber, Mangy Ngo

February 2022

Outline

- Motivation: The Need for Enterprise-Level Agile Data and Analysis Capabilities in MLS Environments
- Challenging Issues for Data Integration and Sharing in MLS Environments
- Proposed Solution: MLS Data Gateway
- Use Cases for Enterprise Data Integration in MLS Environments
  - *Use Case A: One-Way Data Sharing among Multiple Security Domains*
  - *Use Case B: Data Integration Via Data Virtualization Products*
  - *Use Case C: Data Integration Via Data Semantics Layer*
  - *Use Case D: Data Access Using Standard SQL-Based Data Access and Query*
- Summary

Motivation: The Need for Enterprise-Level Agile Data and Analysis Capabilities in MLS Environments

- Data management in space systems needs to evolve
  - *From data silos with system-specific, pre-determined (waterfall-like) data management*
  - *To enterprise-level, agile data integration and analysis capabilities*
- Next-generation space data-centric architectures must also support agile and effective data sharing and integration in Multi-Level Security (MLS) environments
  - *MLS environments: Data reside at different levels of security, and policy-based data dissemination among security levels is required*
- A common scenario in the space enterprise: A user in a lower-level security enclave needs access to data residing in a higher-level security enclave
- The MLS environment should:
  - *Allow the user to issue generic queries for the necessary data*
  - *Enforce policies on:*
    - What queries are allowed to pass from the low side to the high side
    - What query results are allowed to pass from higher to lower security enclaves

Characteristics of Enterprise Data

- Enterprise data is distributed
  - *Data are owned and best managed by various organizations (systems)*
  - *Legacy systems need to be easily integrated into enterprise level data management*
- Emergent use cases require on-demand data analytics
  - *Run enterprise data analytics on data from diverse resources managed by various organizations*
- **Enterprise and its data tenants are in multiple security enclaves**
  - *Support the ability to access and integrate data from multiple enclaves*
- Enterprise data is dynamic
  - *The velocity of data flowing from data sources (e.g., sensors) is constantly changing*
- Enterprise Data is diverse
  - *Enterprise data may be structured, semi-structured, unstructured, or streaming data*
  - *Enterprise data solutions must blend all these data types with machine-learning data to gain holistic views and actionable insights and to optimize mission operations*

Challenging Issues for Data Integration and Sharing in MLS Environments

- Multiple data access and data sharing protocols make on-demand, programmatic-based data protection more difficult
- In current systems, data sharing among different MLS enclaves has used custom rather than general-purpose solutions
- Until recently, there has been little support for automatic, policy-based and generic data protection in MLS environments
- There is a lack of mature technologies and products for automated and timely sharing of data in MLS environments
- Commercial off-the-shelf (COTS) MLS Guard tools:
  - *A hardware or software Guard enables or restricts the access or transfer of data between security domains based on a predetermined security policy*
  - *Guards for MLS environments don't provide high-level interfaces*
    - Typically offer limited and lower level data access protocols (e.g., Extensible Markup Language (XML))
  - *The latest generation of Guard products provide programmatic data protection assertions and offer support for a few(but limited number of) higher level protocols*

A General Approach to Alleviate the Challenges of Data Sharing in MLS Environments

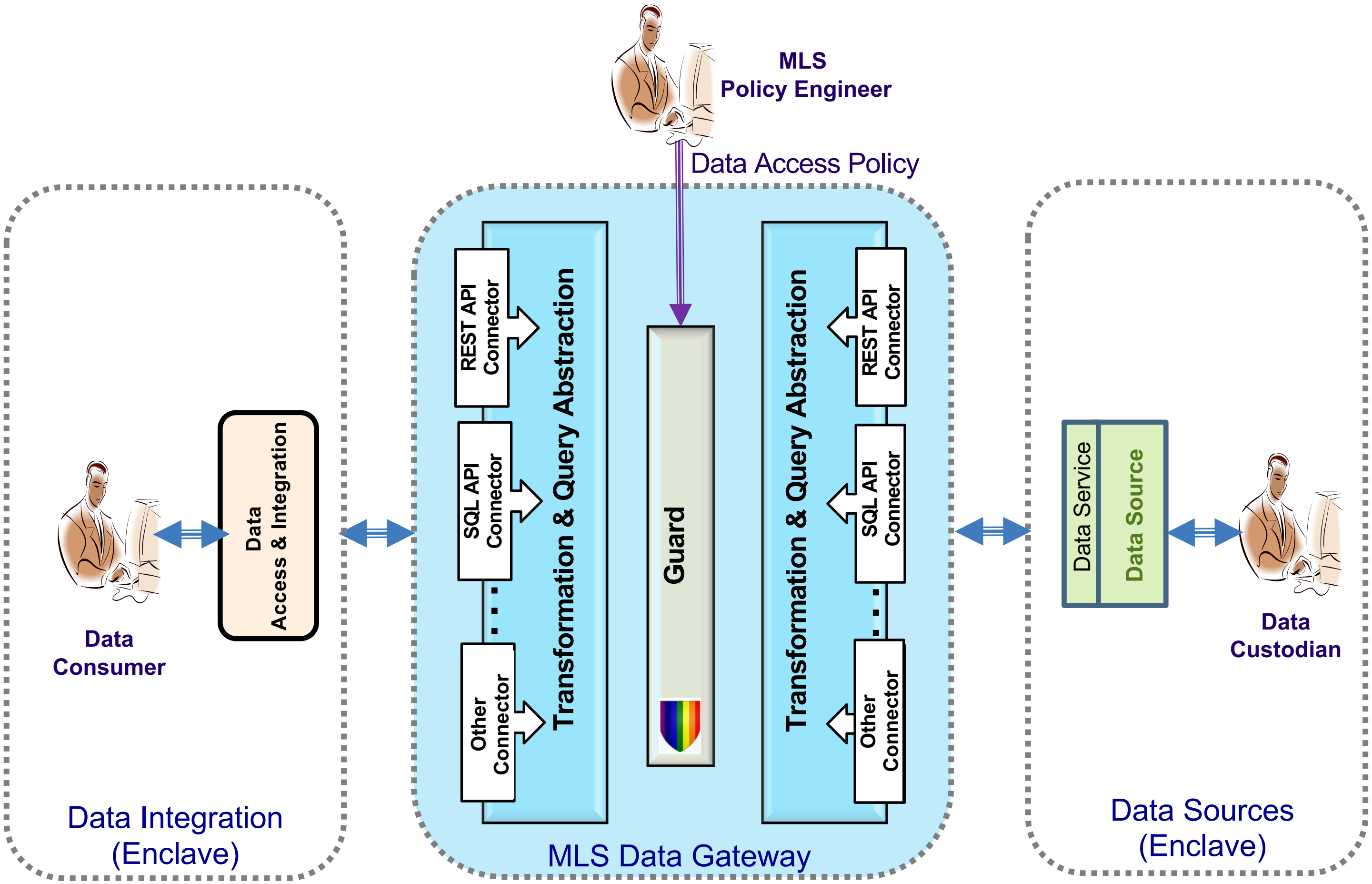
Solution should include several elements:

- Should use a **uniform approach for data integration and sharing** in both single security domains as well as in multi-level security domains
- Should **extend the use of data integration tools and techniques that were developed for a single enclave to multiple enclaves**, with the tools being MLS-aware
- Should support **policy-driven data protection** that provides automatic (whenever possible) data sharing among enclaves with the required level of data protection
- Should help **abstract a physical data source** and its specific characteristics
- Should provide **fine and coarse granularity data access control**
  - *Access control for data should be at both (i) the macro level (on the data set), as well as (ii) the more granular data item (data entity) level*

Proposed Solution: MLS Data Gateway

- A **Data Gateway** is a software and hardware solution that:
  - *Connects to multiple data systems*
  - *Provides a single, central point of access to safely access data with the necessary data protection*
- A Data Gateway acts as a bridge that provides secure data transfer between environments or enclaves
- An **MLS Data Gateway** supports agile and effective data sharing and analysis in multi-level security (MLS) environments, where:
  - *Data reside at different levels of security*
  - *Policy-based data dissemination among security levels is required*

Anatomy of an MLS Data Gateway



Basic Characteristics of an MLS Data Gateway

An MLS Data Gateway includes:

- **Guard:** specialized hardware and/or software that provides a controlled interface to **enable or restrict the access or transfer** of data between **two or more security domains** based on a predetermined **security policy**

- **Connectors** for standard data access protocols:
  - REST API
  - SQL queries

- **Transformation and Query Abstraction** component:
  - ***Translate*** *from the incoming data access protocol to the protocol required by the Guard (e.g., XSD (XML Schema Definition))*
  - *Incoming request (e.g., a data access request from a consumer in Enclave 1 for data stored in Enclave 2) is transparently transformed into the format required by gateway*
  - *Then transformed again into the format required by the other enclave*

- MLS Gateway can also provide data filtering capabilities
  - *Allowing data filtering at a much higher level of granularity*

Use Cases for Enterprise Data Integration in MLS Environments

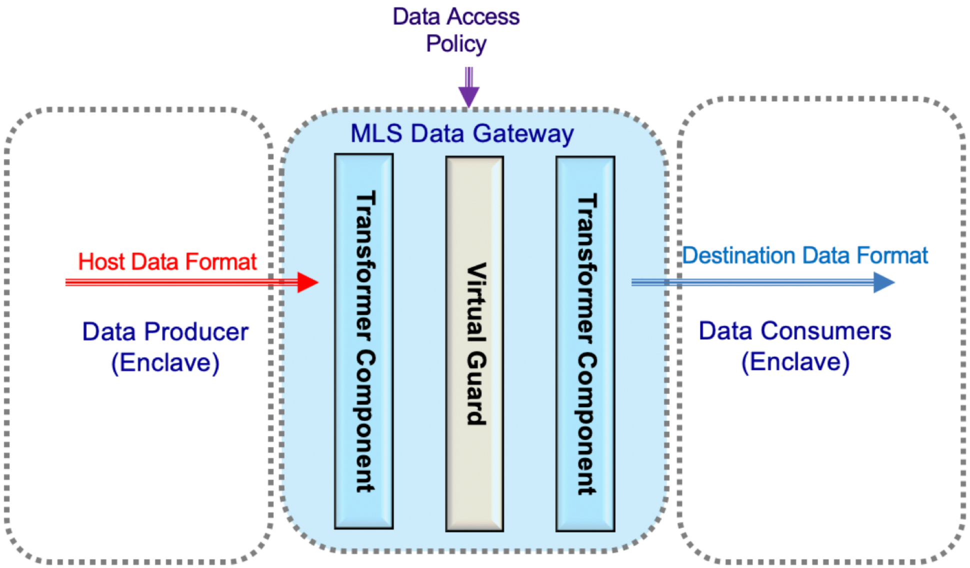
- Use Case A: One-Way Data Sharing among Multiple Security Domains
  - *Programmatic transfer of data from one security domain to another with policy-based data protection*

- Use Case B: Data Integration Via Data Virtualization Products
  - *Support for enterprise data integration with data sources residing in different security domains using COTS data integration products*

- Use Case C: Data Integration Via Data Semantics Layer
  - *Support for data sematic layer-based integration for data fusion, with the data sources residing in multiple security domains*

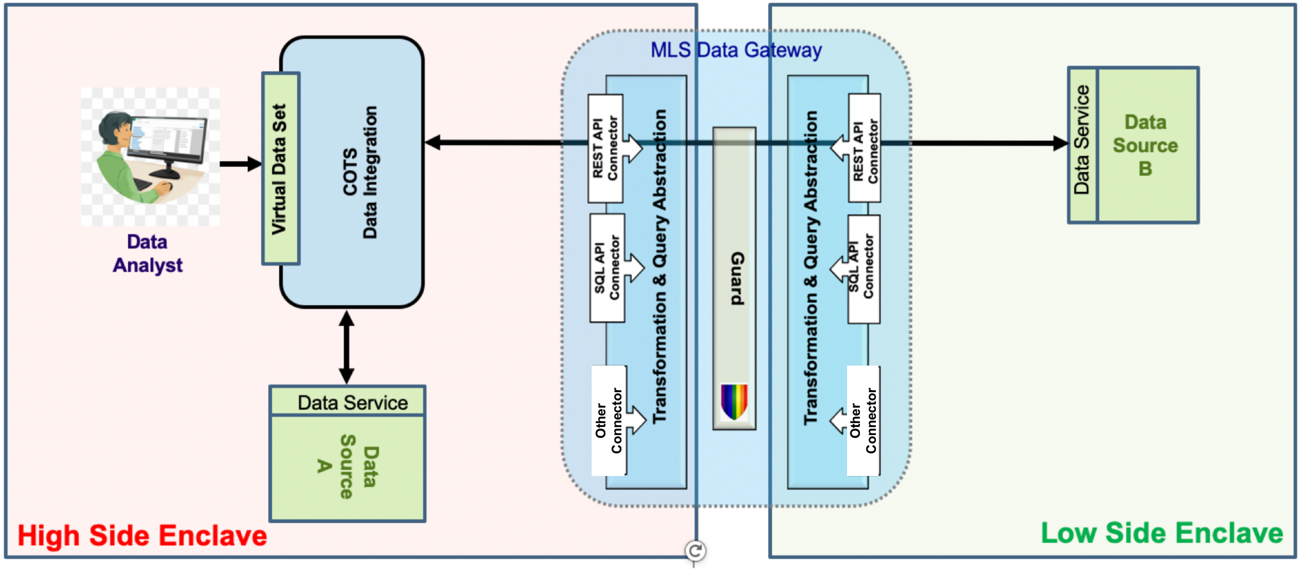
- Use Case D: Data Access Using Standard SQL-Based Data Access and Query
  - *Support for data sharing using standard SQL queries when the data requester and data sources are in different security domains*

Use Case A: One-Way Data Sharing in MLS Environments



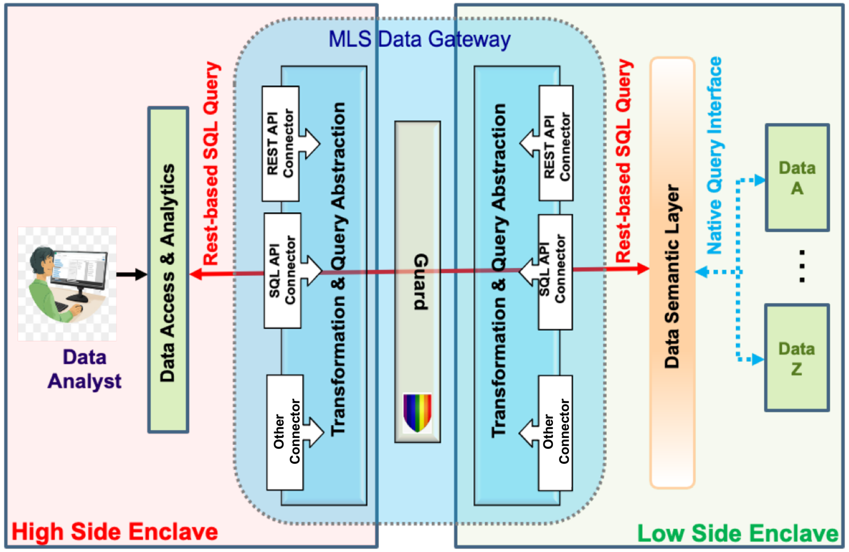
- Data may be shared with multiple consumers residing in multiple security domains
- Data access policies may be based on both data content as well as data tags and are enforced by an MLS Guard
- Current hardware or virtual MLS Guard technologies and tools provide data protection support for a small set of data formats (most commonly XML)
- Our solution encapsulates the MLS Guard with the Data Transformer component
  - The Transformer component is responsible for transforming the data from the host data format to XML (and vise versa)
- The combination of Transformer and MLS Guard (denoted in this project as the **MLS Data Gateway**) allows **data sharing with defined access policies** between enclaves with different levels of security

Use Case B: Data Integration Via Data Virtualization



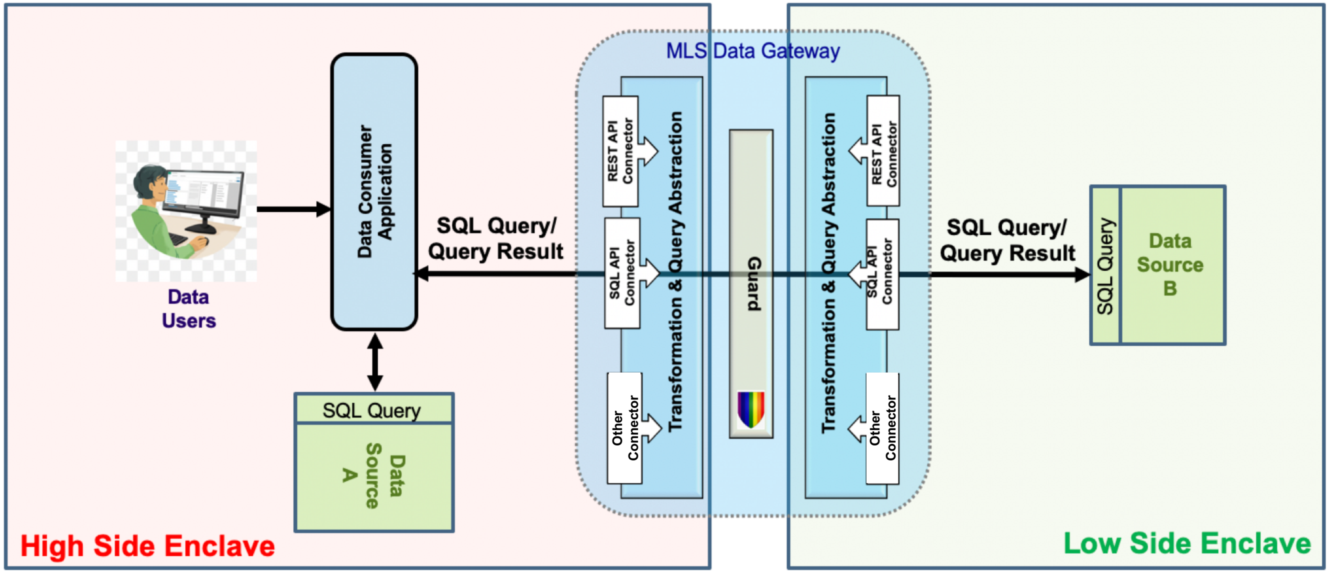
- Enterprise level data integration, where the data sources from multiple systems can be correlated and fused, is a critical element of enterprise data analysis and management
- Emerging COTS data virtualization products are leading technologies for data integration, but they only support integration for data sources that reside at the same security level
- Integration of COTS data virtualization products with the MLS Data Gateway would allow programmatic and safe data integration and virtualization in MLS environments
- Use case: The data analyst discovers available data sources located locally or in other security enclaves and then accesses and integrates those sources using the COTS data virtualization tool
- The Data Gateway provides policy-based access to the remote data source(s) that reside in other security domains

Use Case C: Data Integration Via Data Semantics Layer



- A semantic layer is a unified representation of various data sources that helps data consumers access data autonomously using common data access interfaces
  - *Provides level of abstraction for data consumers that hides fragmentation of data sources*
  - *Unfortunately, COTS data semantics products support data integration for data sources that reside at the same security level*
- The integration of COTS Data Semantic Layer products with the MLS Data Gateway would support programmatic and safe semantic-based data integration in MLS environments
- Use Case: The data consumer accesses the Semantic Layer via REST-based query requests; the endpoint of the Data Semantics Layer is managed by the Data Gateway
- The Data Semantic Layer provides semantic-based access to diverse data sources
- MLS Data Gateway is responsible for (i) data protection policies (ii) filtering of query results

Use Case D: Data Access Using SQL-Based Queries



- SQL-based queries to retrieve raw or processed data are most popular data access protocol
  - *Unfortunately, there is no support for SQL queries in COTS Guard products*
- The integration of an SQL-based data consumer application with the MLS Data Gateway would support programmatic and safe data integration in MLS environments
- Use case: Data analyst accesses data via SQL queries when the data requester and data sources reside in different security domains
- The SQL query for Data Source B is intercepted by the SQL Connector/Transformer (part of the MLS Data Gateway) and delivered to the Guard as an XML message
- After passing the Guard, the message is transformed back to SQL format by Transformer
- The SQL message is delivered to the data source, which then returns the query results through the MLS gateway to the data consumer application

Summary

- The new generation of data centric architectures must support agile and effective **data sharing and analysis in multi-level security (MLS) environments**, where:
  - *Data reside at different levels of security*
  - *Policy-based data dissemination among security levels is required*

- MLS Data Gateway Architecture and prototype have demonstrated the feasibility of secure, policy-based, agile/dynamic methods of data integration among enclaves with different security levels
  - *Completed these prototype implementations of use cases for data sharing:*
    - One Way Data Sharing among multiple security domains
    - Data Integration Via Data Virtualization products
    - Data Integration Via Data Semantics Layer products
    - Data Access Using Standard SQL Based Data Access and Query

- A key insight from these use cases: The MLS Data Gateway **integrates seamlessly with:**
  - ***Existing commercial data integration, virtualization and semantic layer-based tools***
  - ***Standard data sharing and interaction protocols*** *(e.g., SQL queries) in an MLS environment*

- This seamless integration **makes advanced data integration and sharing capabilities that were originally developed for a single security enclave available for use in MLS environments**