Cloudera Runtime 7.3.3

Cloudera Security Overview

Date published: 2020-07-28
Date modified: 2026-12-13



https://docs.cloudera.com/

Legal Notice

© Cloudera Inc. 2026. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Introduction 4
- What is Cloudera on premises? 4
- Importance of a Secure Cluster 4
- Secure by Design 4

Pillars of Security 5
- Authentication 5
- Authorization 5
- Encryption 5
- Identity Management 6

Security Management Model 6

Security Levels 7
- Choosing the Sufficient Security Level for Your Environment 8
- Logical Architecture 8

SDX 9

Security Terms 10

Introduction

Securing data is important to every aspect of your business, including your customers and personnel. Before you can confidently and effectively secure your data, you must understand data security fundamentals.

Cloudera on premises, and the components that make up Cloudera, rely on these fundamentals to ensure your data is secure by design.

This document provides a high-level overview of security basics including the importance of securing a cluster, the Four Pillars of Security, and Cloudera Shared Data Experience, as well as a comprehensive glossary of security terms.

What is Cloudera on premises?

Cloudera on premises is an integrated analytics and data management platform built for hybrid cloud and deployed in on-premises data centers.

It consists of Cloudera Base on premises and Cloudera Data Services on premises, offering broad data analytics and artificial intelligence functionality along with secure user access and data governance features.

Importance of a Secure Cluster

Threats to your cluster can come from a variety of sources, so securing your cluster against all of these sources is important.

External attackers might gain access to your cluster and compromise sensitive data, malicious software might be implemented through known vulnerabilities, and insiders or third party vendors might misuse legitimate permissions if authorization is not appropriately implemented.

Governing bodies have implemented various data protection and privacy laws which you must implement into your cluster's security. For example, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) is a United States federal law that requires protecting sensitive patient health information from being disclosed without the patient's consent or knowledge. The General Data Protection Regulation (GDPR) is a regulation in European Union law aimed at enhancing individuals' control and rights over their personal data both inside and outside the EU.

Additionally, if your business accepts credit card payments, you must comply with the Payment Card Industry Data Security Standard (PCI DSS). This global standard applies a series of security measures to ensure the safety of your customers' credit card information.

It is therefore crucial to have a secure cluster that addresses all potential threats and vulnerabilities, and complies with all data protection and privacy laws. By actively reducing the threat surface on your environment, you actively manage the risk, likelihood, and impact of a security event.

Secure by Design

Our Cloudera is secure by design.

We treat security as the top-tier design requirement to manage risk, reduce attack surface and vulnerabilities, and develop design concepts and patterns in an industry-preferred way.

We have designed Cloudera to meet technical audit requirements out-of-the-box. In fact, many Cloudera services require security components prior to their installation.

Cloudera gives you the ability to manage your cluster securely, allowing you to audit and understand how and what has happened on your cluster, and manage data policy and governance.

Pillars of Security

To ensure the security of your data, Cloudera follows the Pillars of Security: authentication, authorization, encryption, and identity management.

These pillars work together to support our platform.

Authentication

Authentication is a basic security requirement in a digital environment.

Users, processes, and services must verify their identity using credentials known by the authenticator in order to access a system. Credentials can include something you know such as a password or PIN; something you have such as a badge or RSA token; or something you are such as biometrics including fingerprints, facial features, or iris patterns.

The components of Cloudera that provide authentication solutions to your cluster are Apache Knox and Kerberos. Apache Knox provides perimeter security so that the enterprise can extend access to new users while also maintaining compliance with security policies. Kerberos is a network authentication protocol designed to provide strong authentication for client/server applications by using secret-key cryptography.

An iPhone user can unlock their device by using the facial recognition scan or inputting their PIN.

Authorization

Authorization is a basic security requirement in a digital environment.

Authorization controls what users can and cannot access within a system. These permissions include viewing, editing, using, or controlling information. In a physical environment, this can also include accessing servers, rooms, and devices.

Apache Ranger is a component of Cloudera that provides authorization solutions to your cluster. Ranger defines the policies that determine what users can and cannot access.



Note: Authorization through Apache Ranger is just one element of a secure production cluster: Cloudera supports Ranger only when it runs on a cluster where Kerberos is enabled to authenticate users.

The owner of a Google Doc can share their document with others. When sharing, they can set the access permissions of others as either editors, viewers, or commenters to determine their level of interaction with the file.

Encryption

Encryption increases the level of security in a digital environment.

Encryption protects sensitive data through encoding, so the data can only be accessed or decoded with a digital key. Encryption applies to both data at-rest and data in-transit.

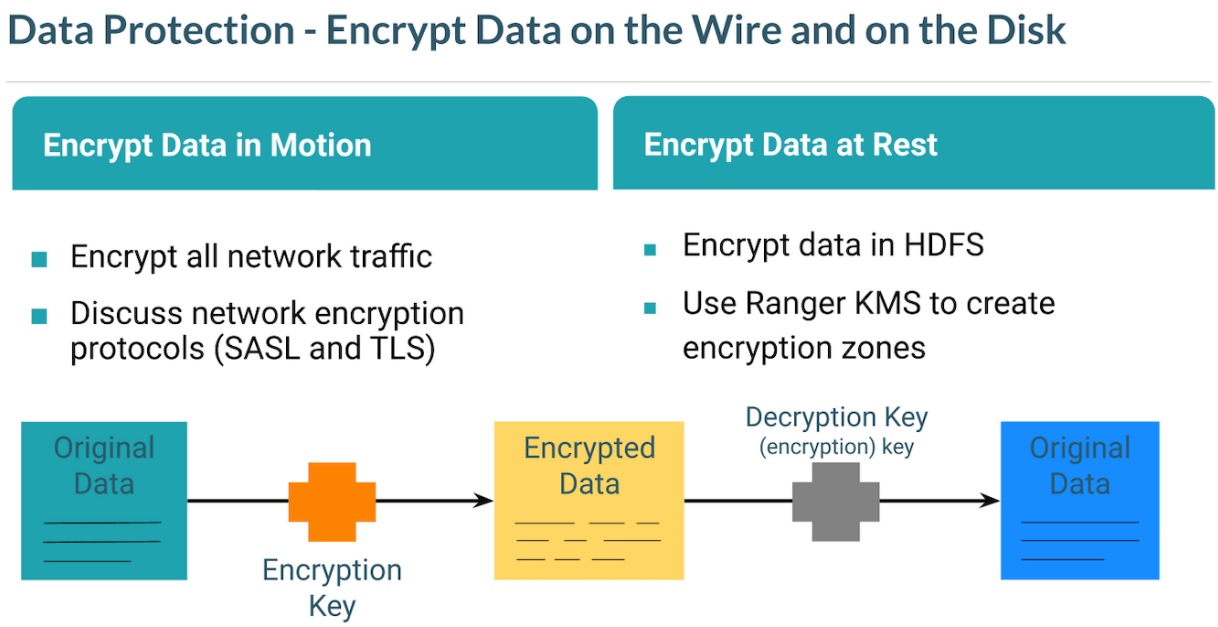
Data at-rest, or data that is stored physically on a storage device such as a hard drive or cloud storage and not actively moving, is encrypted and digital keys are distributed to authorized users to decrypt it when needed.

Data en route between its source and its destination is known as data in-transit. End-to-end encryption utilizes protocols like Transparent Layer Security (TLS) to protect data in-transit. Each data transmitting session has a new digital key, which relies on proper authentication and can reduce the risk of unauthorized access.

Cloudera provides the following components for encryption solutions: Ranger KMS and Navigator Encrypt.

Ranger extends the Hadoop KMS functionality by allowing you to store keys in a secure database. Navigator Encrypt transparently encrypts and secures data at rest without requiring changes to your applications.

Figure 1: Encryption model demonstrating encryption at rest and in transit



Identity Management

Identity management is how user and group identities are centrally stored.

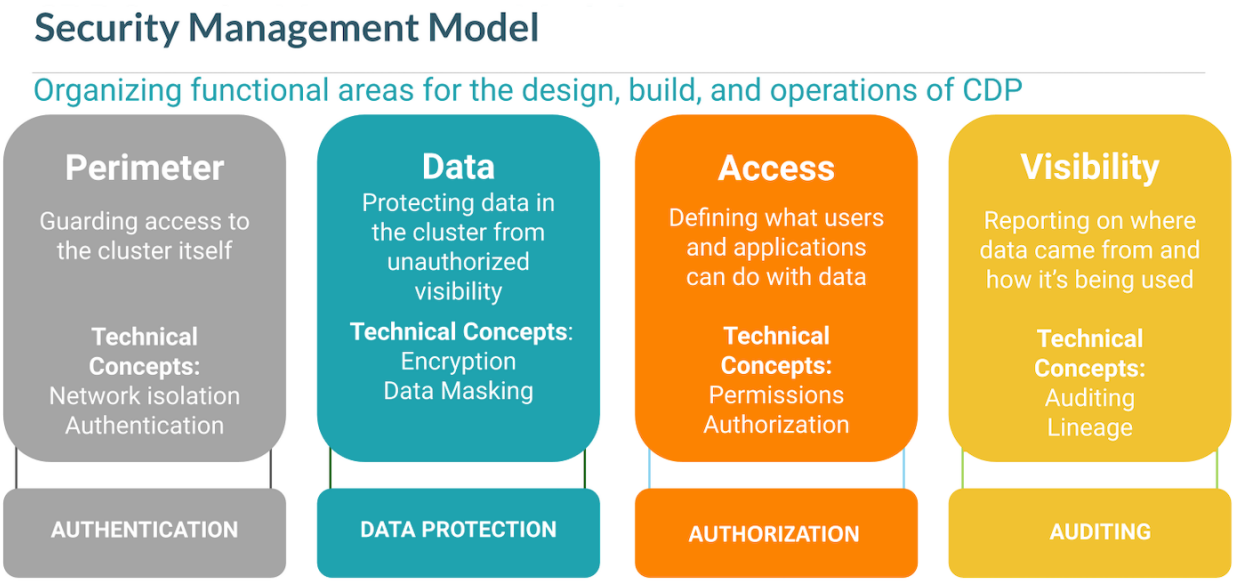
Identity Management functions in conjunction with authentication and authorization as a form of security that ensures everyone has appropriate access to the resources needed to perform their roles. Managing identities of users and groups involves implementing policies and allowing or denying permissions.

Security Management Model

The Pillars of Security combined with related technical concepts and Cloudera components make up the Security Management Model.

The combination of Cloudera Manager, Ranger, and Atlas allows you to consolidate security configurations in your clusters. The key security principles of our layered approach to data defense are as follows:

Figure 2: Security Management Model



Related Information
Security Management
Security Levels

Security Levels

Securing a cluser requires a layered approach, applying the Pillars of Security to ensure data availability, integrity, and confidentiality. Cloudera on premises offers four security levels, as security is no longer optional.

0 - Non-Secure

The cluster has no configured security. Clusters without security must not be used in production environments because they are highly vulnerable to all attacks and exploits. Cloudera does not support running an insecure environment.

1 - Minimally Secure

Minimally secure clusters are configured for authentication, authorization, and auditing. Authentication is first configured to ensure that users and services can only access the cluster after proving their identities. Authorization mechanisms are then applied to assign privileges to users and user groups. Auditing procedures keep track of who accesses the cluster.

2 - More Secure

Clusters that encrypt their sensitive data are more secure. These systems include a variety of steps, including key management, to ensure data is encrypted while in-transit.

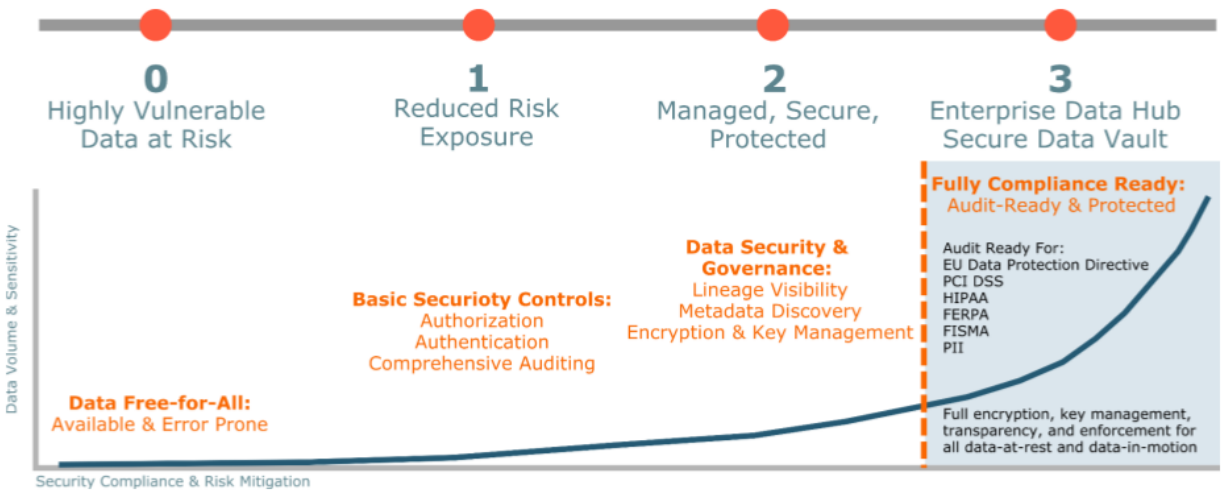
Auditing has been set up for data governance to trace any data object's lineage and verify its authenticity. It determines what people are doing with your data, and when they are doing it.

3 - Most Secure

The most secure clusters ensure that all data, both at-rest and in-transit, is encrypted and the key management system is fault-tolerant. Auditing mechanisms comply with industry, government, and regulatory standards such as PCI,

HIPAA, GDPR, and NIST, and extend from the cluster to any integrated systems. Cluster administrators are well-trained, the security procedures are certified by an expert, and the cluster passes technical reviews.

Figure 3: Cloudera Security Level Model



Choosing the Sufficient Security Level for Your Environment

Cloudera will assist you in deciding which security level is best for your environment.

Regardless of the security level each major organization chooses, they must complete a penetration test, also known as a “pen test.” This test simulates a cyberattack through ethical hacking which evaluates the system security, identifying its strengths and weaknesses.

Related Information

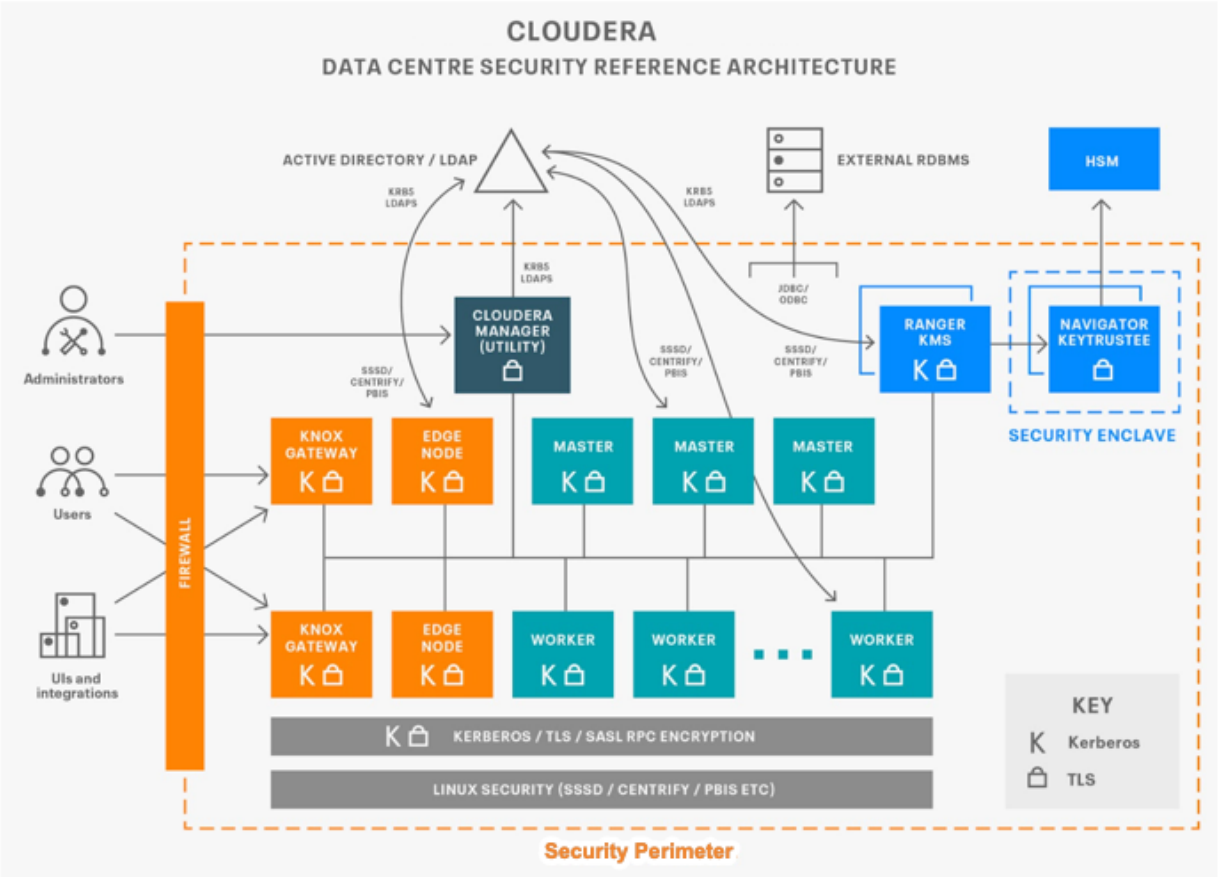
Cloudera Security

Logical Architecture

The components that embody the Pillars of Security and make up the Cloudera on premises security architecture are designed to work together logically.

The following diagram visualizes this relationship:

Figure 4: Cloudera on premises Security Architecture



SDX

Cloudera Shared Data Experience (SDX) ensures your information is secure by design.

What is SDX?

SDX is Cloudera's design architecture that we incorporate into all of our products. It ensures your information is secure by design, so an integrated set of security and governance technologies protect your data. These technologies are built using metadata, which SDX collects to implement security policies.

SDX contains a combination of tools such as Ranger, Atlas, Knox, Hive Metastore, Cloudera Data Catalog, Cloudera Replication Manager, and Workload Manager.

SDX and Security

SDX provides consistent policy, schema, and metadata on the backend. It is effectively a single pane of glass for viewing metadata, schema, and policy within your digital environment to reduce security risks and operational costs.

Related Information

Cloudera SDX: Cloudera Shared Data Experience
Apache Ranger
Apache Atlas
Apache Knox
Apache Hive Metastore
Cloudera Data Catalog

Cloudera Replication Manager in Cloudera Base on premises

Security Terms

A glossary of commonly used data security terms.

Access Management

A framework for controlling what users can and cannot access.

Access Control List (ACL)

A list of associated permissions that specifies which users or system processes are granted access to a digital environment.

Apache

The Apache Software Foundation is an American nonprofit corporation that supports open source software projects. Cloudera utilizes Apache software.

Apache Atlas

Apache Atlas is a scalable and extensible set of core functional governance services that provides open metadata management for organizations. Atlas enables enterprises to effectively and efficiently meet their compliance requirements within Cloudera.

Apache Hadoop

The Apache Hadoop software library is a scalable framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models.

Apache Knox

Apache Knox provides perimeter security so that the enterprise can confidently extend Hadoop access to new users while also maintaining compliance with enterprise security policies. Knox also simplifies Hadoop security for users who access the cluster data and execute jobs.

Apache Ranger

Apache Ranger is a framework to enable, monitor, and manage comprehensive data security across the platform. It is used for creating and managing policies to access data and other related objects for all services in the Cloudera stack.

Auditing

The process of assessing the quality of data to ensure its authenticity and integrity.

Authentication

The process of proving an individual is who they claim to be.

Authorization

The protection that allows or denies user access to certain resources through specific established rules.

Certificate

Digital certificates are small data files that digitally bind a cryptographic key to an organization's details. It is used in electronic documents to verify the identity of an individual, a server, or an organization.

Data At-Rest

Data that is not actively moving and is stored in a physical location such as a hard drive or a cloud.

Data In-Transit

Data that is actively moving from one location to another, whether across the internet or between devices.

Data Integrity

Maintaining and assuring the accuracy and validity of data.

Data Governance

The prcess of managing the availability, usability, integriy, and security of data based on established policies.

Data Lineage

Lineage informatin helps you understand the origin of your data and the transformations it may have gone through before arriving in a file or table.

Data Masking

Protect sensitive data from being viewed by unauthorized users through methods such as redacting, hasing, nullifying, or implementing a partial mask.

Data Protection

Prevent the accidental deletion of data files through various protective methods.

Encryption

A security method where information is translated from plaintext to ciphertext and can only be accessed or decrypted by a user with the correct encryption key. Encrypted data appears as unreadable to anyone accessing it without permission.

Kerberos

Kerberos is a network authentication protocol. It is designed to provide strong authentication for client/server applications by using secret-key cryptography.

Key

Within cryptography, a key is needed to encrypt and decrypt a message. Keys determine the output of the cipher algorithm.

Key Management

The management of cryptographic keys in a cryptosystem. This includes dealing with the generation, exchange, storage, use, destruction, and replacement of keys.

Identity Management

A framework of policies that verifies user identities and ensures only authorized users have access to certain data or technology.

Lightweight Directory Access Protocol (LDAP)

Lightweight Directory Access Protocol (LDAP) is a software protocol for enabling anyone to locate organizations, individuals, and other resources such as files and devices in a network, whether on the public Internet or on a corporate intranet.

Metadata

Data about data.

Permissions

The authorization that enables users to access specific resources such as files, applications, and devices.

Personal Identifiable Information (PII)

Personal Identifiable Information (PII) includes any information that could be used to identify an individual.

Policies

A set of guiding principles or rules established to enforce data governance which determines how data is collected, used, and accessed.

Secure by Design

Cloudera treats security as a tier-1 design requirement to manage risk, reduce attack surface and vulnerabilities, and develop design concepts and patterns in an industry-preferred way

Transport Layer Security (TLS)

Transport Layer Security (TLS) is the most widely used security protocol for wire encryption. TLS provides authentication, privacy, and data integrity between applications communicating over a network by encrypting the packets transmitted between endpoints.