TMS 5.1 OTP Planning Guide

Version 2

May 2010

Date of Publication: May 2010

Last update: May 2010

Support

We work closely with our reseller partners to offer the best worldwide technical support services. Your reseller is the first line of support when you have questions about products and services. However, if you require additional assistance you can contact us directly at:

Telephone

You can call our help-desk 24 hours a day, seven days a week:

USA: 1-800-545-6608

International: +1-410-931-7520

Email

You can send a question to the technical support team at the following email address:

support@safenet-inc.com

Website

You can submit a question through the SafeNet Support portal:

http://c3.safenet-inc.com/secure.asp

eToken
YOUR KEY TO eSECURITY

Table of Contents

Chapter 1

Introduction

This guide describes the issues involved in planning and designing the eToken One Time Password (OTP) solution, and provides relevant information and recommendations.

In this chapter:

Overview

This guide assists in planning the setup and use of the SafeNet One Time Password (OTP) solution in a network infrastructure. The guide is intended for information technology professionals responsible for the organization's network security.

The information provided in this guide provides the following:

- Information to formulate a plan for an OTP Server installation servicing a small to large number of users.
- Information required to successfully design and implement an OTP solution in an environment based on the Token Management System (TMS) 5.1 server.
- An explanation on the concepts and components of an OTP and TMS 5.1 design.
- Outlines the solution requirements and deployment scenarios.

For a detailed reference guide to TMS 5.1 and the other infrastructure components involved in the solution design or other SafeNet products mentioned in this guide, refer to the product specific documentation.

For additional information on Microsoft and other 3rd party software and hardware components mentioned in this guide, refer to the relevant manufacturers' documentation.

Chapter 2

OTP Concepts

This chapter defines the eToken one-time password (OTP) solution concepts used in this guides explanations.

In this chapter:

About OTP

The OTP authentication method is a multi-factor authentication method replacing static passwords with dynamically changing passwords, making it more difficult to steal or guess a user's password and gain unauthorized access to restricted resources.

OTP authentication relies on an OTP authenticator which is an OTP generation token that generates the next password to be used, displays it for the user and let the user type it in, instead of (or in addition to) their password.

In this way, attackers trying to impersonate the user or penetrate a computer system would require access to the OTP authenticator in order to gain access to the network.

The solution is based on a shared secret (also known as the OTP seed) that is embedded in the OTP authenticator and is also known to the Authentication Server. Since both parties know this OTP seed, at authentication time the authentication server can validate the correctness of the one-time password provided by the user.

The OTP generator modifies the effective password based either on elapsed time (for example, every minute) or on events (for example, every time the user clicks a button on the OTP authenticator).

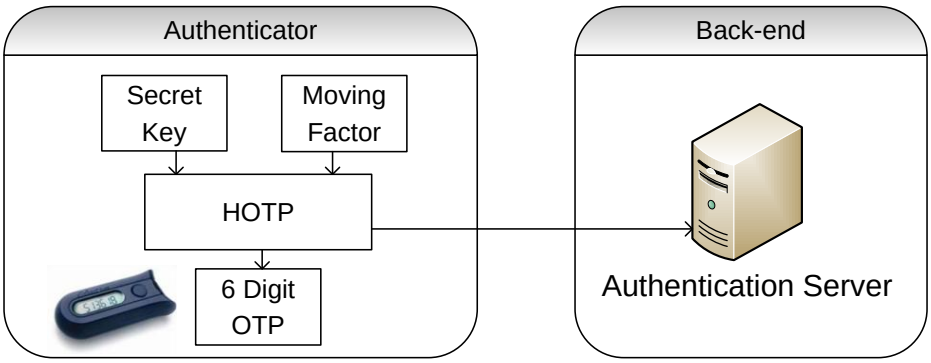
OTP authentication can use one of the following industry-standard or proprietary algorithms for the one-time password calculation and generation from the OTP seed:

- **HOTP:** An HMAC-based OTP algorithm, initially created by the Open AuTHentication (OATH) organization and approved (IETF RFC 4226) in 2005.
- **TOTP**: A time-based OTP algorithm which has been submitted to IETF by the OATH organization, and is in the process of standardization.
- **X9.9**: A legacy banking standard designed and implemented for financial transaction authentication.
- **Challenge Response:** Challenge Response Algorithm. This solution is based on OTP challenge response mechanism that requiring a response from the token based on a challenge provided by the OTP backend.
- Vendor specific/proprietary algorithms.

Additional usable algorithms include the following:

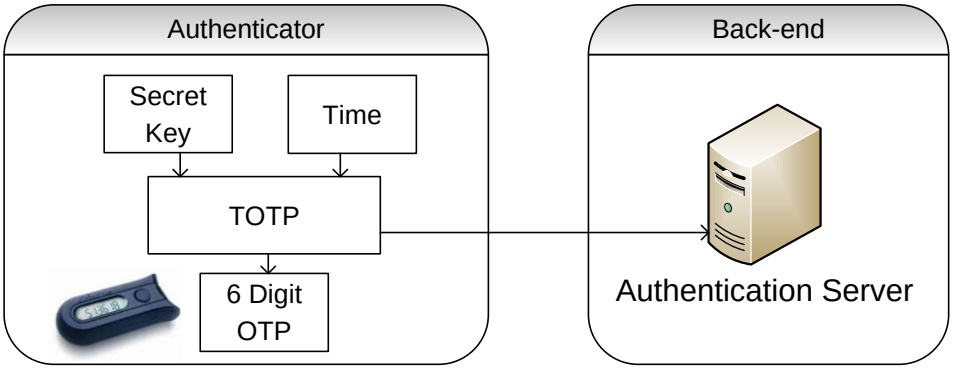
- **OTP PIN:** This solution requires a PIN entry in addition to the OTP value during authentication.
- **PIN Protected:** Used in hardware authenticators. This solution requires a PIN entry on the authenticator itself for it to generate an OTP.
- **Application PIN:** Used in software authenticators. Requires a PIN entry to the application to generate an OTP value.

The following figure illustrates a 6-digit OTP generation using an eTokenPASS physical event.



A new authentication password is generated every time the OTP token is physically clicked.

The following figure illustrates a 6-digit OTP generation using an eTokenPASS time event, for example every 30 seconds.



OTP Solution Usage

OTP authentication generates a password that is only valid for a single or time-limited login session or transaction which assists in avoiding replay attacks. Examples of common OTP authentication uses are described in the following paragraphs.

Mobile Workers

Unlike a user physically located on an organizations premises who is connecting to network resources from a secured perimeter, mobile users connect to organization resources from external networks. There cannot be any reliance on physical security, and using a simple password to validate users accessing the organizational network is insufficient. Using OTP tokens at authentication time makes it much more difficult for attackers, and makes it possible for mobile users to securely access the organization network.

When using OTP authentication, an OTP password generated by the OTP authenticator in addition to their password is required. This combined action strengthens authentication and enabling secure remote access to the organization network using technologies such as VPN.

In Perimeter Security

OTP security is not necessarily restricted to remote access and mobile users. OTP security can be added to an internal network increasing organizational security.

OTP can be added to secure network logon access, or other internal applications used in the organization, such as CRM, SharePoint portal, web applications and terminal servers, currently using a user name and password for authentication.

OTP access may be required for all users or just for a selected user community. An example of such an implementation may be the use of OTP for elevated privileged users and system administrators requiring strong authentication for accessing sensitive infrastructure, applications or resources.

Authentication for Online Services

Online services are web-based services offered over the internet to consumers and business users. Examples of online services include online banking, e-commerce, e-learning or access to patient records (health care). OTP-based strong authentication enables users to securely access confidential information requiring a password generated by an OTP authenticator.

Transaction Validation

OTP can be used to validate and approve specific transactions after logging onto an online web site using basic password authentication.

For example, in an online banking environment, the user may be able to log on to the web site using the user name and password for viewing information, but would require an OTP re-authentication when performing a transaction or funds transfer.

OTP Authenticators

For secure authentication implementation the basic requirements is a personal OTP authenticator available for each user. This section summarizes authentication tokens and modes supported by eToken products.

OTP is an authentication method based on a token and the backend authentication service sharing a common OTP seed used to generate a sequence of different passwords. Because the OTP seed is shared, the server can validate the user-provided OTP by generating the expected password on the server side and comparing it to the generating user-provided password.

There are two categories of OTP authenticators:

- Hardware authenticators are hand-held passcode generators programmed with the same unique cryptographic algorithm used by the authentication server. The authenticator has a Liquid Crystal Display (LCD) to display their generated passcodes and a button to generate a passcode.
- Software authenticators implement the same functionality and logic in software and can be installed on PC computers or mobile phones. The software authenticators generate the OTP which can be displayed on the host computer or mobile phone.

Hardware Authenticators

The following hardware tokens are supported by the Token Management System (TMS):

- **eTokenPASS**: An OTP token, providing user authentication to network resources. The token uses the OATH standard of HOTP or TOTP to generate the one-time passwords. The shared OTP seed is used to generate OTP is embedded in the device at the time of manufacture.

- **eToken NG-OTP**: An integrated hybrid USB smartcard and OTP token. The authenticators operate as either a smartcard when attached to a PC using the USB connector or as an OTP token when detached.

  Like the eTokenPASS, the eToken NG-OTP uses the OATH standard of HOTP to generate the one-time passwords. However, unlike the eTokenPASS, and because of the USB interface, the OTP seed used for OTP generation is not embedded during manufacturing, but is provided dynamically on-site when the token is provisioned using TMS.

  eToken NG-OTP can be ordered with embedded proximity coils, enabling it as an integrated secure logical and a physical access authenticator. Proximity coils are a type of antenna enabling the token device to be used, for example, for physical access to buildings. The proximity coils operate with a corresponding reader installed where the reader permits access only if the matching proximety coil is detected.

- **Gold**: The Gold is designed in a key fob case design. The Gold operates in either synchronous or asynchronous mode, and incorporates additional features including the option of a one-time passcode mode, a pre-expired PIN mode (forcing change of default PIN at first use), the option to set a fixed number of PIN uses, and more user-friendly display prompts.

- **Platinum**: The Platinum tokens have the same features as the Gold, and offer replaceable dual batteries (warranted for 5 years) and providing an unlimited operation. The durable case and housing enables Platinum tokens to have the longest warranty available in the industry.

Software Authenticators

Software authenticators, also known as eToken MobilePASS, are software applications enabling OTP generation on an existing PC or mobile phone without the physical dedicated token requirement.

The following software tokens are supported by TMS:

- **eToken MobilePASS for Blackberry**: Generates an OTP on a RIM Blackberry device.
- **eToken MobilePASS for Windows Mobile**: Generates an OTP on a Windows Mobile powered mobile or smart phone.
- **eToken MobilePASS for Windows J2ME**: Generates an OTP on a J2ME enabled mobile phone.
- **eToken MobilePASS for Windows**: Generates an OTP on the Windows desktop.

OTP Validation

OTP validation is TMS Authentication Server validation of OTP authentication requests. TMS calculates the expected OTP value for a user and compares the value to the information provided by the authentication request.

The user authentication application, whether it is a web application, a VPN gateway or another authentication point, must be integrated with TMS to validate incoming users. The application integration can be by applying various standard protocols, TMS OTP authentication plug-ins, or direct application integration using the TMS OTP SDK.

TMS

eToken TMS connects users, their security devices, and the organizational policies to the associated security applications. eToken TMS links them all into a single automated and fully configurable system, enabling the implementation of enterprise-wide token management services.

In an OTP implementation, TMS performs two major roles:

- **Authentication validation server:** TMS validates incoming authentication requests and returns a response with the authentication result (accept or deny) to the requesting service, such as a VPN gateway. For more information, see OTP Integration on page 11.
- **OTP life cycle management:** TMS provides the services enabling administrators, helpdesk, and end-users to manage and deploy tokens. For more information, see OTP Life Cycle Management on page 12.

New in TMS 5.1:

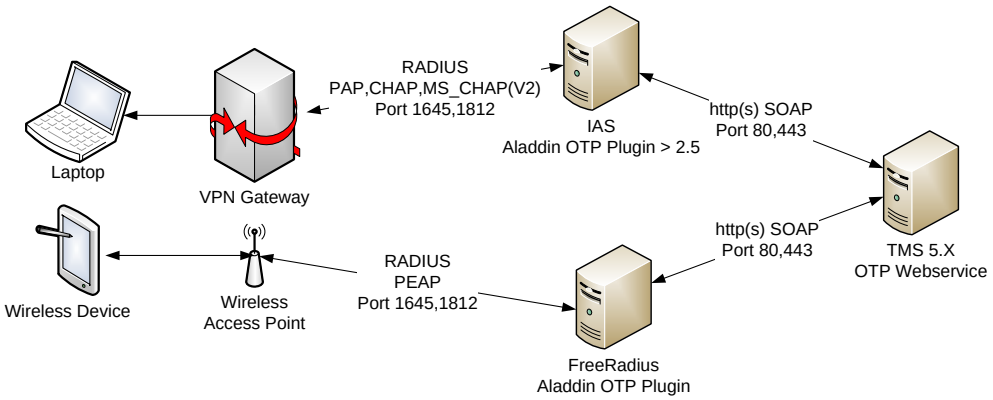
- SafeWord 2008, SafeWord Premier Access, SafeWord Remote Access servers can be migrated to TMS 5.1.
- Support for SafeWord tokens. TMS supports the following OTP tokens: SafeWord Alpine, SafeWord Gold and SafeWord Platinum.
- Support for time based OTP authenticators: eTokenPASS Time.
- Automatic synchronization for out of sync OTP tokens.

OTP Integration

Applications and network services requiring OTP authentication and enforcement must be integrated with TMS authentication services.

TMS provides out-of-the box integration modules, and supports standard base protocols for ease of integration. For full flexibility and extendibility, TMS also offers the OTP integration SDK for application developers.

The following diagram illustrates the components involved in a typical OTP environment and the associated data flow. It illustrates how an end-user can use a VPN gateway together with RADIUS servers to validate incoming authentication requests.



The OTP authentication flow is as follows:

1. A client requires access to a VPN gateway or WLAN access point.
2. The access gateway prompts the user for authentication credentials including username, PIN and the OTP value.

3. The service uses these credentials to authenticate the user via the RADIUS protocol, handing over the request to the RADIUS server (IAS server or FreeRADIUS).
4. The eToken OTP authentication plug-in installed on the RADIUS server validates the request via web services (SOAP over HTTPS) to the TMS validation service.

In the following sections it is assumed that Microsoft IAS and eToken IAS plug-in are used. However, the methods presented can also be applied to the FreeRADIUS server and eToken FreeRADIUS plug-in configuration.

OTP Life Cycle Management

eToken TMS capabilities include token deployment and revocation, web-based user self-service token enrollment and password reset, and handling of lost and damaged tokens.

For example, the following can be performed:

- If a token is lost, the Help Desk or the end-user can replace the token with a new one; or the end-user can create a temporary static password instead of the OTP until the token is found.
- A permanent static password can be used with the OTP when performing authentication, so that two-factor authentication can be enabled (this is called the OTP-PIN).
- Using the TMS helpdesk specific functions can be performed, such as resetting the OTP-PIN or synchronizing the token, for example if the user has generated too many OTPs on the token without authentication.

Overall, the services provided by TMS are grouped in three distinct categories:

- **Management services:** for administrators and Help Desk staff.
- **End-users self service:** for end-users self help and management.
- **Remote access and recovery service:** for remote end-users, including recovery options.

Chapter 3

Before You Begin

This chapter analyzes specific issues considered when planning and designing the integration of OTP in an organization, and provides detailed information about the available options and their implications.

In this chapter:

Determining the Usage Scenario

When beginning the planning process, ensure to have the specifications available about the current network environment. Specifically, the hardware and software inventory and a network topology map which can be very helpful in reducing time spent during the design phase.

It is important to know the users, their location, and the network resources they will access with OTP.

Identify the Users and Their Physical Locations

The first step is to identify the users' location. Check if all users are located in one physical site or in multiple physical sites, and if users are also connecting remotely.

It is recommended to answer the following questions at this point:

- How many users are there on the main site?
- Are there more organization sites, if so how many?
- How many users exist on each site?
- How many users are mobile users connecting remotely?
- How many users connect when traveling out of the office?

The location of users, how frequently they require access and the usage profile are important aspects of the performance and availability design of the deployment. It is also important to list all the different usages of OTP in the environment, and to know which gateways the users will connect to, such as an SSL VPN Gateway, or the organization's Outlook Web Access (OWA).

User Roles

Four main user roles need to be assigned, as described as follows:

- TMS System Administration
- TMS Security Administration
- Help Desk Team
- End-users

TMS System Administration

The TMS System Administrator role is responsible for installing the system, applying patches and fixes, managing the backup of the system, and ensuring its high availability.

The number of TMS System Administrators may vary, depending on the size of the organization and other internal considerations.

The TMS System Administrator main roles are as follows:

- Responsible for installing the system, preparing the hardware, installing the Windows Server operating system, network connections, disk drives, etc.
- TMS Server maintenance, including applying operating system patches, TMS software patches if necessary, and performing regular system backups.
- Estimating the number of simultaneous authentications per second, the number of users and physical sites existing in the organization, and implementing the solutions accordingly. For a sizing and performance matrix, see *Chapter 3* Authenticators Comparison Matrix.
- Answering questions, such as “How many servers need to be installed?” and “Is it necessary to install servers also on other sites?”.
- Ensuring the availability of machines and the system. The TMS system administrator deals with high availability and scalability planning of the TMS infrastructure. For more information, see *Chapter 4,* Architecture.

TMS Security Administration

The TMS Security Administrator governs organizational security policies, such as configuration, defining TPOs, assigning roles to the Help Desk, etc.

The TMS Security Administrator defines who can use the system, best security practices, and integration with applications that require security.

The number of Security Administrators may vary, depending on the size of the organization and other internal considerations.

The TMS Security Administrator main roles are as follows:

- Reviewing organization security policies and identifying sensitive resources requiring additional strong authentication.
- Configuring the system according to the organization security policy for example, setting the minimum OTP PIN to be 6 digits instead of 4. (The OTP PIN is the fixed password used with the OTP).
- Creating logistics policies, for example, answering the following questions: “How should the tokens be distributed to the users?” and “Will the Help Desk pre-enroll the tokens to the end-users or will the users enroll the tokens for themselves?”.
- Help Desk team and end-users education on how to use the TMS and manage the tokens. This can be done for example by using booklets, documentation or with on-site training.
- Delegating OTP token life cycle management to the TMS Help Desk team. Using the TMS Authorization Manager, the TMS Security Administrator can restrict Help Desk users performing life cycle management for only some of the users, based on specific OUs or Active Directory groups (if the TMS is configured with AD as the user repository).

  For example, a Help Desk user named *helpdesk1* could be allowed to manage the token lifecycle of users in an OU named *Marketing*, but not for other users.

  In addition, the Help Desk users can be restricted to only some of the operations. For example, they could be given the permission to enroll tokens, but not to reset the OTP-PIN.
- Delegating OTP token life cycle management to certain end-users based on AD OUs or groups, allowing them to perform only some of the self-service options, such as enrolling a token but not generating a temporary OTP.
- Assigning the Help Desk team its responsibilities and delegated authorization.

Help Desk Team

The TMS Help Desk is responsible for OTP token life cycle management.

The number of Help Desk users can vary depending on the organization, and can contain users from one site or multiple sites. For example, the Help Desk team can contain five users from the main site and one or two users from each remote site.

The TMS Help Desk team main roles are as follows:

- Prepare an end-user token physically (or via carriers) give it to the user, or will can an unassigned token to the end-user. The end-users can enroll the token, according to the logistics policy defined by the administrator.
- After giving the tokens to the end-users, the Help Desk team is responsible for managing the token. For example, if a user has forgotten the OTP PIN, the Help Desk team can reset the OTP PIN.
- According to the security policy of the organization, when a user leaves the organization (in most cases), the user returns the token to the Help Desk team, which un-assigns the token from the user and removes it from the database.

  The token is then empty and can be used as an authentication device for any other user.

End-users

End-users are users who are actually using the tokens for authentication. They can receive already pre-enrolled tokens from the Help Desk team, or they can receive new blank tokens which they can enroll for themselves.

The end-users are assigned their permissions by the TMS Security Administrator.

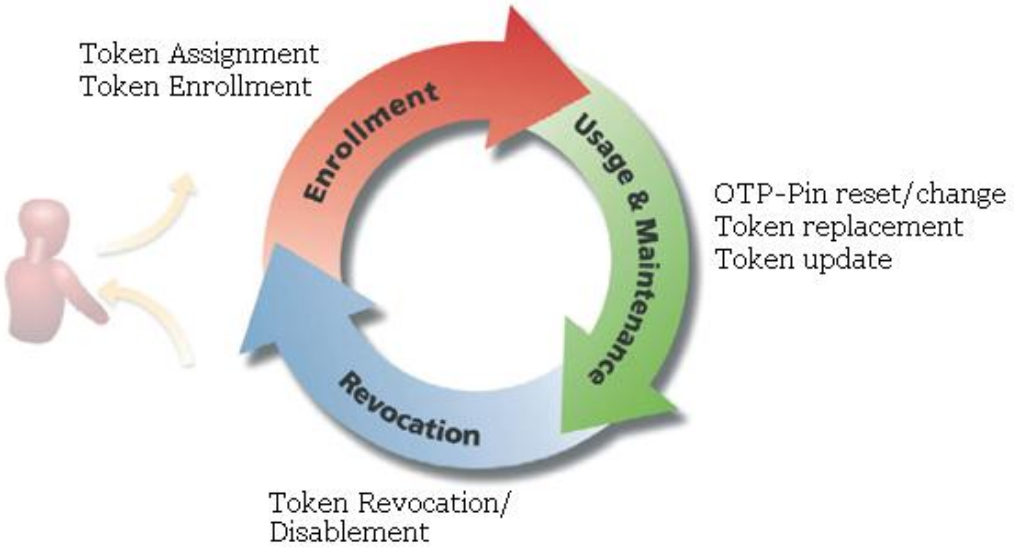
Some of the TMS end-users capabilities are as follows:

- Given a pre-enrolled token, the end-user can use the OTP device to connect to the secured resource.
- Given a blank token, the end-users can assign or enroll the token for themselves before using the token for authentication.
- If the end-user has any issues with the token, there is a self service portal to help manage the token, for example, to reset his OTP PIN.

- Once the end-user has finished using the token, the token is returned to the Help Desk team, as planned in the administration policy (in most scenarios).

Understanding Token Lifecycle Management

The following diagram illustrates the token life cycle, from assignment/enrollment, to usage and management, revocation, and the assignment/enrollment to another user.



To address life cycle management, eToken TMS capabilities include:

- Token deployment and revocation
- Web-based user self-service token enrollment and password reset
- Handling of lost and damaged tokens

Distribution and Assignment

Assigning/enrolling a token for a user means that the token is assigned to a specific user. Whenever the user tries to login to an OTP secured gateway, the token is required for authentication.

Distribution and assignment of tokens to users can be performed in two ways:

- By the Help Desk or Administrator users, who prepare the tokens for the end-users and then distribute them, physically or by mail.
- Using the Self-Service TMS web site, which enables end-users to enroll their own tokens.

The enrollment can be combined, so that some of the users receive the already prepared tokens, and other users enroll tokens for themselves.

For considerations on which method to choose, see *Selecting the Distribution Method* on page 30.

Temporarily Disabling an eToken

For security reasons an enrolled can be temporarily disabled. This temporary disablement can be for long periods of time.

OTP PIN Reset

The OTP PIN is an additional fixed password used together with the changing OTP value to increase security and to enable secure two-factor authentication (The OTP PIN is “something you know”, and the OTP value is “something you have”).

Using TMS the OTP PIN can be reset if the end-user has forgotten the PIN.

OTP Synchronization

If a user presses the OTP token many times without authenticating (too many blank presses), the token can become unsynchronized and the user will not be able to authenticate. The number of blank presses is configurable.

Automatic OTP Sync, which is new in TMS 5.1, keeps the OTP seed in sync with the server, resetting it automatically whenever necessary.

Supports both time and event based OTP tokens.

Note: For more information, refer to the *eToken TMS 5.1 Administrator's Guide*.

Temporary OTP

If an OTP token is lost or damaged, the TMS can enable a temporary (Temp) OTP to replace the OTP token function.

A Temp OTP is a static value to use in place of a generated OTP. Its value does not change, and so it provides only a low level of security. It is valid for a limited time until the user finds his authenticator or obtains a replacement.

Lost Token

For security reasons, it is important that a lost or damaged token is revoked as soon as possible.

Note: Depending on the TMS configuration, when a user is deleted from the AD domain, the user's tokens are automatically unassigned.

When a token is revoked, the following occurs:

- The token's status is set to *Revoked* in the TMS inventory.
- The token remains associated with its user.

Lock/Unlock OTP

Temporarily locking an OTP disables its use for OTP authentication. Locking an OTP token may be useful if a user temporarily loses or forgets the token's location, or in periods of inactivity of the user, for example while on vacation etc.

Unlocking the OTP re-enables it for use for OTP authentication.

Understanding Self Service

eToken users have access to the TMS Self Service Center web site and are able to manage their tokens using the facilities on the site.

Self Service Operations

The TMS configuration specifies policies determining which token activities end-users are authorized to perform.

eToken TMS is designed to help ensure that no one except the authorized user uses the eToken.

End-users can be authorized to perform the following activities:

- Enroll/update the eToken:
  - Enroll the eToken
  - Update the eToken
  - Upgrade the token by replacing it with a new authenticator
  - Replace a lost token with a new one
- Maintain the OTP token using a TMS service center:
  - Temporarily disable the token if it is misplaced or if it is not needed for an extended period
  - Enable the disabled token when the user needs to use it again
  - Synchronize the OTP if OTP values were repeatedly generated without being submitted for authentication, and the OTP token has lost its synchronization with the system
  - Reset the OTP PIN

For more information on TMS operations, refer to the *eToken TMS 5.1 Administrator's Guide*.

Before opening the TMS self-service portal to end-users, it is important to consider the following:

- Are all the end-users sufficiently skilled enough in order to use the TMS self-service portal?
  - What kind of training the end-users needs to be given to start working with the TMS self-service portal?
  - What documentation, booklets, or training is needed to be supplied to some or all of the users?
  - Will the users be able to enroll the tokens for themselves, or is the Help Desk team required to do the enrolling?
  - If some of the users have accessibility issues, Help Desk enrollment might be required. It is important to identify the users that may have issues with self-enrollment.
  - Can the end-users manage their own tokens, or will they need assistance from the Help Desk?
  - Can the end-users get assistance from the Help Desk when they need support?
- For which end-users is the self-service portal accessible?
  - Is it open to all end-users or only to some of them?
  - From which networks will the users be able to access the TMS self-service web site? From their LAN, WAN, extranet, wireless network?
- Which functions will all end-users or groups of end-users be able to perform?
  - Will end-users be able to assign/enroll tokens for themselves?
  - Will end-users be able to replace their token?
  - Will end-users be able to reset the OTP PIN?

Security for Self Service

TMS supports an advanced delegated administration enabling different administrators to have control of different users and groups.

For example, it is possible to define that a specific OU named *Marketing* is able to enroll a token but is not be able to reset the OTP PIN, and another group named *SalesPersons* is allowed to enroll their token and reset the OTP PIN.

Determining Integration Requirements

It is important to list all the different network gateways that need to be integrated with OTP, in addition to the existing servers in the organizations that hold the user repository.

Authentication Clients

It is important to identify the specific services used in the organization requiring strong two-factor authentication based on OTP.

These services are OTP authentication clients and need to integrate with the TMS OTP validation service.

For example, an authentication client can be one of the following:

- VPN Gateways, for example, Cisco Concentrator, Checkpoint Firewall-1.
- VPN SSL Gateways, for example, Cisco ASA, Checkpoint Connectra, Juniper.
- Web servers, such as Microsoft IIS.
- Microsoft Outlook Web Access.
- Citrix Web Interface.
- Proprietary applications. Integration with these may require the use of the eToken OTP SDK. For example, the usage of the OTP SDK in Oracle database, added OTP support for strong authentication.
- For more information on the eToken OTP SDK, see Selecting the Integration Technology on page 45.

Databases and User Repositories

TMS stores, manages and maintains OTP token information in its Token repository, including the token status, the OTP seed used to generate the OTP and the token assignment to users. For user information, TMS is designed to be integrated with an external user store. During the design process it is important to identify which user repository the organization is using, such as Microsoft Active Directory.

If the customer is not using any external user store, TMS 5.1 uses an internal user store created and maintained by the TMS server itself.

eToken TMS 5.1 supports the following external user repositories by default:

- Microsoft Active Directory – 2003 and 2008
- Novell eDirectory
- Open LDAP
- Microsoft SQL Server 2005 and 2008

It is important to know which user repository is being used in the environment, and as well as the repository structure. It is recommended to check and complete the following questionnaire:

- General – Applicable to all user repositories:
  - In which networks are the servers located?
  - Is there a firewall between the servers and the TMS servers?
  - How many users exist in the user repository?
  - What is the version of the product (for example, Microsoft Windows 2003 AD or Microsoft SQL 2005)?
  - Which Service Packs are installed?
  - If more than one server exists:
    - Which type of replication is performed between the servers?
    - What is the replication interval between the servers?
- For Microsoft Active Directory:
  - How many forests are used?
  - How many domains are used in each forest?
  - How many domain controllers are used in each domain?
  - How many physical sites are used, and what is the network bandwidth between them?
- For Novell eDirectory:
  - How many trees are used in the environment?
  - How many NDS servers in each tree?
  - How many physical sites are used, and what is the network bandwidth between them?

- For Open LDAP:
  - How many Open LDAP servers are used?
  - Description of the Open LDAP schema.
  - The authentication method that is used in OpenLDAP.
- For Microsoft SQL:
  - How many SQL servers are being used?
  - Description of the SQL schema – users tables etc.
  - Which authentication method is used? For example, the passwords are hashed using SHA-1.
  - How is the connection to the SQL database implemented, directly or using ODBC?

Choosing Authenticators

This section describes and compares the different authenticators.

Authenticators Comparison Matrix

Hardware authenticators available for use with TMS 5.1 are eTokenPASS, eToken NG-OTP, Gold, Platinum.

The software authenticators, also known as MobilePASS, are a software OTP solution available for Windows desktop, and mobile phones: Windows Mobile, Blackberry, Symbian, and JavaME enabled mobile phones.

It is important to decide when to use hardware authenticators, when to use software authenticators and when to mix and match them.

The following table lists the main considerations for authenticator type selection:

	Hardware Token	Software Token
Security	More secure: End-users use a physical device to generate an OTP.	Less secure: Although the end-users cannot copy the software OTP file to another machine, it is still considered as less secure than a hardware device.
Security - Hardware PIN Protected or Challenge response	More secure: Users can use tokens with Protected PIN or with OTP Challenge response for higher security.	Less secure: In Software authenticator there is no option to use Protected PIN or OTP Challenge Response.
Price	More expensive.	Less expensive.
Physical distribution	More complex: End-users obtain replacement tokens by physical token shipment and pickup.	Less complex: End-users obtain the token via electronic delivery and self-service.
Assign-ment/enrollment	Less complex: Help Desk and end-users assign/enroll the hardware device.	More complex: Help Desk and end-users install an application in Windows or on a mobile phone.
Using the token	Less complex: End-users click on the device to generate an OTP.	More complex: End-users open an application (either on Windows or on a mobile phone) to generate an OTP.
Carrying the token	More complex: End-users carry another hardware device which may get lost or misplaced.	Less complex: End-users do not need to carry additional devices and can use an application on an existing device (Windows OS or mobile phone).

The following table lists the main considerations for hardware authenticator selection:

	eTPass	NG-OTP	Gold	Platinum
Synchronous	Yes	Yes	Yes	Yes
Event Based OTP	Yes	Yes	Yes	Yes
Time Based (TOTP OATH)	Yes	No	No	No
Asynchronous (Challenge Response)	No	No	Yes	Yes
Hardware PIN Protected	No	No	Yes	Yes
OATH Compatible	Yes	Yes	No	No
Enrollment	Seed file	USB Enrolment	Seed file	Seed file
Battery Replacement	No	No	No	Yes

Difference between Enrolling eTokenPASS and eToken NG-OTP

When enrolling the eTokenPASS token, which is not USB-based, the enrollment is done without connecting the device to the end-user or the Help Desk enrollment client. The end-user or Help Desk user enters into the TMS the eTokenPASS serial number (written on the back of the token), and the TMS enrolls it in the database.

When enrolling the eToken NG-OTP, the end-user or Help Desk user is connects the token to the enrollment client, and the TMS enrolls the token for OTP usage.

In both scenarios, an additional password can be added to the OTP (either a fixed OTP PIN or a Windows password) during the enrollment process.

For more information on the enrollment process, refer to the *eToken TMS 5.1 Administrator's Guide*.

User Validation and Token Distribution

This section describes the different options available to validate and distribute hardware and software tokens to the end-users.

When the tokens arrive at the organization in which the OTP solution is being implemented, the tokens are assigned or enrolled to users.

There are two main options for enrollment:

- Helpdesk (on-behalf) enrollment
- Self-service enrollment

Help Desk (on-behalf) Enrollment

- Hardware:
  - Keep the tokens at the IT Help Desk department, which is responsible for assigning and enrolling the tokens for the users in the organization. End-users can physically go to the Help Desk team, and ask them to enroll a token for them. The Help Desk team personally identifies the user and prepares a token for the user.
  - When using Help Desk enrollment, the end-users receive the prepared token and do not need to go to the TMS server themselves to enroll a token. This reduces the Help Desk workload.
- Software:
  - The end-users physically go to the Help Desk team with their mobile phone or flash device, and can ask them to enroll a token for them.
  - The Help Desk team personally identifies the end-users, and can connect the end-user's mobile phone or flash device to their machine and create a software OTP profile on the flash device or mobile phone.

Self-service Enrollment:

- Hardware:
  - Distribute blank/unassigned tokens to the end-users (either distribute them physically or ask the users to take them from the Help Desk team themselves). The users will then be able to connect to the TMS server with their network credentials and assign/enroll a token for themselves.
  - When using self-service enrollment, the workload of the Help Desk team reduces significantly.
  - It is also possible to create a self-enrollment station in a designated place in which users can log in, insert their token and enroll themselves an OTP token.
  - When distributing tokens to multiple sites, it is possible to send packages of tokens to the physical sites and have the users self-enroll their tokens. Another option is to give a few key personnel the option to enroll the tokens for part/all the users at the site.
- Software:
  - End-users can connect to the TMS self service portal and download a software OTP profile to their machine.
  - If the end-users are using a Windows application, the end-users need to download the profile to a folder on the machine, and run the application whenever they need to generate an OTP.
  - If the end-users are using a mobile phone, the end-users need to connect their mobile phone to the machine and copy the application to the mobile phone. In this case, the OTP seed is stored inside the application.

The advantage of TMS is that is allows a full matrix of options. For example, administrators can have some of the users enrolling the tokens themselves, others will get an already prepared token, and some can use a self-service station to enroll their own token.

Selecting the Distribution Method

The following table lists the main considerations for distribution selection:

	Help Desk - Hardware	Help Desk - Software	Self service - Hardware	Self service - Software
Enrollment and maintenance	Easier for end-users. Requires resources from Help Desk.	Easier for end-users. Requires resources from Help Desk	More complex for end-users. Easier for Help Desk.	More complex for end-users. Easier for Help Desk.
Enrollment process	Easy: Assign/ enroll token.	More complex: Requires connection to mobile phone or installing Windows application.	Easy: Assign/ enroll token.	More complex: Requires connection to mobile phone or installing Windows application.
Physical distribution	Need to distribute hardware devices.	No additional device needed.	Need to distribute hardware devices.	No additional device needed.

Token Replacement and End-of-life

OTP tokens work on battery power, which lasts between 3-7 years before the battery runs out and the token stops functioning.

As a result, the existing tokens need to be replaced with new tokens after a certain period of time.

There are two main options when setting up token replacement logistics:

- Replacing the token before the battery runs out. This is the recommended option, since it ensures that the users can continue to use their tokens without interruption.
- Replacing the token when the battery runs out. This requires the user to call the Help Desk to receive a new token, which may result in hours or days during which the user will not be able to authenticate using the OTP token.

It is possible to mix the two options - some of the users could receive new tokens near the expiration of their tokens, and others when their tokens expire.

Chapter 4

Architecture

This chapter describes the TMS solution components, explains the principal architecture options, and provides information and recommendations to assist with deployment.

In this chapter:

Choosing an OTP Engine

The following chapter focuses on TMS installation as the OTP backend engine. The TMS provides all the required components as described on the next section.

An alternative option for using the TMS, is to use the SafeNet OTP Authentication Engine SDK solution.

About SafeNet OTP Authentication Engine

The SafeNet OTP Authentication Engine is a standalone set of APIs and sample code that integrates with an organization's environment and provides the backend functionality required for deploying SafeNet One-Time Password (OTP) authentication solutions, without having to install a separate full-scale backend management server. The SafeNet OTP Authentication Engine enables organizations to leverage their existing environments and infrastructure to roll out large-scale OTP deployments where high performance and throughput are required.

The API includes functions for:

- Importing OTP token records from import files provided by SafeNet
- Validation for event/time base OTP authentication requests
- OTP token management operations

The SafeNet OTP Authentication Engine is designed to address the needs of organizations and enterprises implementing OTP authentication for large scale online Web-based services such as online banking, ecommerce, e-learning, and access to online health records. The SafeNet OTP Authentication Engine addresses the following specific needs:

- Organizations requiring to implement large-scale OTP deployments using their existing infrastructure and environments
- Organizations that need high performance and throughput such as B2C environments – banking, ecommerce, payment portals
- Organizations that do not require comprehensive token management capabilities or already have these capabilities in existing platforms
- OTP environments that require high performance and throughput

- Self contained environments where a second authentication server is not required
- Organizations that require support for additional platforms include the following:
  - Windows Server 2003 SP2 x32/x64
  - Windows Server 2008 x32/x64, Windows Server 2008 R2 x64
  - Fedora 11
  - Ubuntu 9.04
  - Centos 5.3. x32/x64
  - RedHat 5.3 x32/x64
  - Suse 11.1

For more information on SafeNet OTP Authentication Engine please refer to your local distributer.

Components of a TMS Solution

This section describes the TMS solution architecture components.

Token Management Server

The Token Management Server provides the following sites:

- **Management site**: Used by TMS administrators and Help Desk for token enrollment and life cycle management.
- **Self-service site**: Used by end-users for self-service token management.
- **Remote self-service site**: Used by employees not at the organization's location ("on the road") as a rescue web site to manage cases of lost tokens or forgotten passwords.

RADIUS Servers

RADIUS protocol is used for authentication and authorization. The eToken OTP solution supports the following RADIUS servers:

- **Microsoft IAS/NPS:** The Microsoft IAS service (used in Windows 2003) and Microsoft NPS service (used in Windows 2008) which is a Windows service running a RADIUS server. This service may be extended by adding

plug-ins for the authentication process. eToken extends IAS for this purpose, such as RADIUS authentication requests which are not verified using the default network password but using OTP validation.

- **FreeRADIUS:** FreeRADIUS is an open source based RADIUS server providing additional platform and protocol support. For more information, see http://www.freeRADIUS.org/.

OTP Clients

Any application requiring the use of OTP requires the following:

1. Obtain the OTP value and PIN from the user.
2. Validate the values with the TMS authentication service.

There are three key possibilities for integrating with TMS validation services:

- RADIUS
- OTP authentication plug-in
- TMS OTP SDK

The RADIUS clients communicating with the RADIUS servers include third party network access control applications, such as VPN or RAS.

eToken OTP can virtually work on any device that has RADIUS implementation for user authentication.

Some examples of OTP clients that can be used with eToken OTP are as follows:

- **Routing and Remote Access (RRAS):** RRAS is a network service in Microsoft Windows Server 2000, 2003, and 2008 providing the following services:
  - Dial-up remote access server
  - Virtual private network (VPN) remote access server
  - Internet Protocol (IP) router for connecting private network subnets
  - Network Address Translator (NAT) for connecting private networks to the Internet
  - Dial-up and VPN site-to-site demand-dial router

- **Virtual Private Network (VPN) Gateways:** A VPN is a private network that uses a public network (usually the Internet) to connect remote sites or users together. Instead of using a dedicated connection such as leased line, a VPN uses virtual connections routed through the Internet from the company's private network to the remote site or employee.
- **eToken OTP Plug-ins:** There are network access applications provided by SafeNet, which control access for end-user applications by verifying credentials with the RADIUS server, providing OTP access to these applications. These include the following:
  - Internet Server Application Programming Interface (ISAPI) filter to control access to Outlook Web Access (OWA) – both OWA 2003 and OWA 2007 (New in TMS 5.1)
  - ISAPI filter to control access to general IIS virtual directories
  - Access control plug-in to Citrix Web Interface (WI)

Note: In Citrix WI 5.0 and higher versions, RADIUS support is built in and does not require an additional plug-in.

OTP 5.1 and later versions support redundancy and failover by enabling the connection to several RADIUS servers. If one RADIUS server fails, another one can take over.

ISAPI Filters

The Internet Server Application Programming Interface (ISAPI) is a multi-tier Internet Information Services (IIS) API.

ISAPI consists of two components, extensions and filters. ISAPI extensions are true applications running on IIS. ISAPI filters are implemented as DLLs loaded into a process controlled by IIS.

ISAPI filters modify or enhance the functionality provided by IIS. The filters always run on an IIS server and filter every request until locating a request requiring processing.

IIS and Virtual Directory

Microsoft Internet Information Services (IIS) is a set of Internet-based services for servers using Microsoft Windows.

IIS includes support for the website creation, configuration, and management.

Each Internet services can publish from multiple directories. Each directory can be located on a local drive or across the network by specifying the directory with a Universal Naming Convention (UNC) name, a username and password to use for access permission. A virtual server can have one home directory and unlimited other publishing directories. These other publishing directories are referred to as virtual directories.

Citrix Web Interface (WI)

Citrix Web Interface (WI) accesses Citrix MetaFrame Presentation Server applications and content through a standard web browser.

The web interface dynamically creates an HTML representation of server farms for Citrix Presentation Server sites.

Direct Application Integration using OTP SDK

The eToken TMS OTP SDK (5.1) provides an API for using eToken OTP services.

The eToken TMS OTP SDK enables components to be added to an application that sends an authentication request for user validation directly to the TMS Server.

In addition the SDK gives the option to integrate some lifecycle management operations to an existing or new web application.

For more information on the eToken TMS OTP SDK, see Selecting the Integration Technology on page 45.

User Store

The user store is not specific to eToken TMS and is usually present before the eToken TMS installation, acting as the users' directory for the organization. To avoid duplication of information and to simplify synchronization, eToken TMS does not duplicate the existing user store to an additional store, but uses it directly.

eToken TMS requires only a read access to the user store, as eToken TMS specific data is kept separately. This ensures that the integrity of the original organizational data is not compromised.

The users' directory is usually a hierarchical database with special notation for identifying objects and their attributes, as well as a query language for finding and enumerating specific objects or a set of objects. This language is called Lightweight Directory Access Protocol (LDAP).

eToken TMS can be used with the following:

- **Microsoft Active Directory** (Windows Server 2003 or 2008): Active Directory (AD) multi-forest user stores are supported. eToken TMS 5.1 supports AD running on Windows Server 2008 as user store and configuration store.
- **Microsoft OpenLDAP:** OpenLDAP is a free, open source, multi-platform implementation of LDAP. It includes scalability, replication and referral features, and supports security features such as ACLs, SSL/TLS/SASL, MD5 and SHA.
- **Novell eDirectory**: Novell eDirectory is an X.500 compatible directory service software product for centrally managing access to resources on multiple servers and computers within a given network. Novell eDirectory is a hierarchical, object oriented database representing all the assets in an organization in a logical tree. Assets can include people, positions, servers, workstations, applications, printers, services, groups, and so on. Novell eDirectory supports partitioning at any point in the tree and replication of that partition to any number of servers.

  eDirectory supports referential integrity, multi-master replication and has a modular authentication architecture. It can be accessed via LDAP, DSML, SOAP, ODBC, JDBC, JNDI and ADSI.
- **Microsoft SQL Server:** Microsoft SQL Server is a Relational Database Management System (RDBMS). It creates computer databases for the Microsoft Windows family of server operating systems.

- **Internal User Store:** The use of an ADAM directory to include both a Configuration Store and a User Store is supported. This enables TMS to be installed in an "all-in-one" mode, without requiring the pre-installation of configuration store and user store databases. This configuration also enables full compatibility with SafeWord.

  To support the new integrated configuration, a user management console enables the system administrator to modify users, groups, and OUs.

  When working with the integrated configuration, TMS Configuration Wizard can be used to install secondary TMS servers, enhancing TMS solution redundancy and scalability.

Token Configuration Store

While the user store is required only for reading the users' details and keeping a link between users and their token(s), the eToken TMS configuration store is where eToken TMS data, such as configuration, token information, and eToken TMS-related user details are stored. The eToken TMS configuration store requires read and write access by eToken TMS software.

The Microsoft Active Directory Application Mode (ADAM) is a directory service running as a user service and not as a system. ADAM is an LDAP directory service. ADAM can run on servers running Microsoft Windows Server 2003 and Microsoft Windows 2000 Server and also on clients running Microsoft Windows XP Professional.

Note: In Windows Server 2008, ADAM has been replaced by AD LDS. To run ADAM on clients running Windows XP Professional, the latest service packs and hot fixes must be installed.

---

eToken TMS supports eToken TMS configuration storage in AD and ADAM.

When a non-AD user store (such as MS SQL Server, OpenLDAP or Novell eDirectory) is used, ADAM is the only available option for the configuration store. It is recommended using ADAM as the eToken TMS configuration store. This avoids the requirement to extend the AD schema and simplifies the maintenance and backup of eToken TMS data.

ADAM provides data storage and retrieval for directory-enabled applications, without the dependencies that are required for the AD directory service. ADAM provides similar functionality as AD, but it does not require the deployment of domains or domain controllers. Multiple instances of ADAM

can run concurrently on a single computer, with an independently managed schema for each ADAM instance.

In very large domains with different geographical sites, AD has advantages over ADAM because of its replication capabilities and stronger security mechanisms. In other environments, the ADAM installation is simpler, and supports all eToken TMS requirements.

TMS Audit Store

TMS writes information in the Windows Event Log, such as TMS events auditing. The Windows Event Viewer can be used to see the details of TMS administration events, for example, a user that has logged on to the TMS web site, and another user that enrolled a token.

TMS and Network Perimeter Security Considerations

As previously described, TMS provides three web sites:

- TMSManage for helpdesk users
- TMSService for end-users
- TMSRemote for end-users who are out of office

Users who login to the *TMSManage* or *TMSService* web site are required to submit a user name and password as defined on the customers user repository (such as Microsoft AD or eDirectory).

In AD it is also possible to configure the authentication method in Microsoft IIS as *Windows Integrated Authentication*. With this method the user will automatically be logged on to the TMSManage web site without being prompted for credentials.

This is done with to the Kerberos protocol used in Microsoft Windows which is based on a ticket a user receives while being logged on to a window.

Note: It is strongly recommended to configure the published TMS web sites to work in HTTPS protocol in order to increase security. HTTPS will provide encryption between the client machine and the TMS server and in addition, will allow the client to verify the identity of the TMS Server.

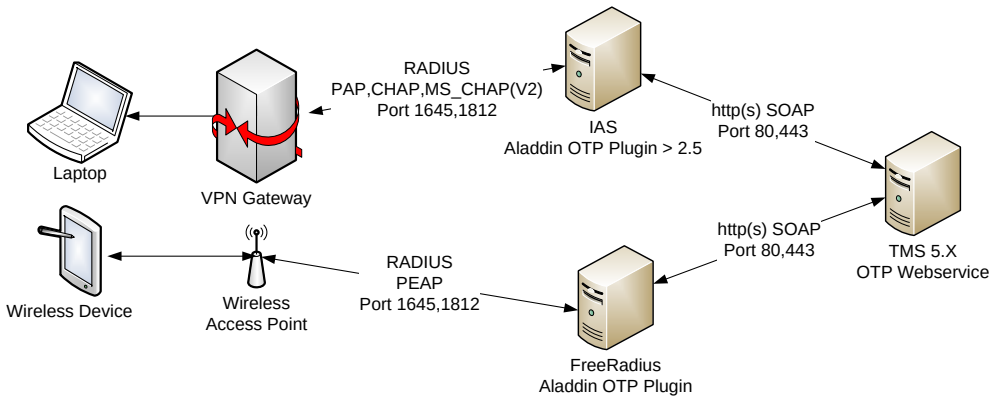
For information on how to configure HTTPS in Microsoft Internet Information Server (IIS) refer to Microsoft documentation.

Firewall and DMZ Considerations

When the TMS is used in a distributed environment such as when the TMS server is on one machine and the IAS (RADIUS) server is on another machine, and there is a firewall between the servers, it is important to note which ports are being used between the servers and to allow access to these ports on the firewall.

If one or more servers are accessible from the outside network or the internet, for example, the VPN Gateway, it is highly recommended to put these servers on the DMZ part of the network, and to open the specific port (for example, 1812 for RADIUS) to the authentication server, which is located in the internal network.

The following figure illustrates an example of a network with VPN Gateway:



The ports that are used and that need to be opened in the firewall(s) are as follows:

From	To	Port
End-user Client machine	Gateway (RADIUS Client)	Depends on the Gateway, for example, for SSL VPN: HTTPS (TCP 443)
Gateway (RADIUS Client)	RADIUS Server	RADIUS (UDP 1645 or UDP 1812)
RADIUS Server	TMS Server	HTTPS or HTTP (TCP 443 or TCP 80)
TMS Server	ADAM as TMS database	ADAM LDAP Port – Default is TCP 50,000
TMS Server	AD (if used as user repository)	53 TCP/ UDP (DNS) 88 UDP (Kerberos)

From	To	Port
		135 TCP (NetBios) 389 TCP / UDP (LDAP) 1025 TCP (Microsoft port) 445 TCP (CIFS) 123 NTP port for synchronizing time 3268 – Global Catalog
TMS Server	MS-SQL (if used as user repository)	MS-SQL (By default , TCP/UDP 1433/1434)
TMS Server	OpenLDAP (if used as user repository)	LDAP (By default, TCP 389)
TMS Server	eDirectory (if used as user repository)	LDAP (By default, TCP 389)

RADIUS Proxy Configuration

Where an IAS server already exists on the network, and the system administrator would like to keep it, an additional IAS server can be added for eToken OTP purposes and to use RADIUS proxy.

RADIUS proxy is where RADIUS clients send authentication requests to a RADIUS server which then proxies the requests to another RADIUS server.

For more information, see *IAS as a RADIUS Proxy* on the Microsoft web site, at: http://technet.microsoft.com/en-us/library/cc785693(WS.10).aspx

Migration from SafeWord 2008 to TMS 5.1

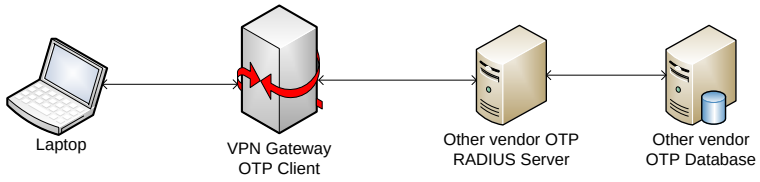
Where the SafeWord 2008 for OTP solution is already implemented, the OTP solution can be migrated to TMS 5.1. TMS 5.1 offers additional solutions such as PKI support.

For more information on how to migrate from SafeWord to TMS 5.1, see *SafeWord migration to TMS 5.1 How to Guide*.

Migration from Other OTP Solutions – Side by Side Deployment

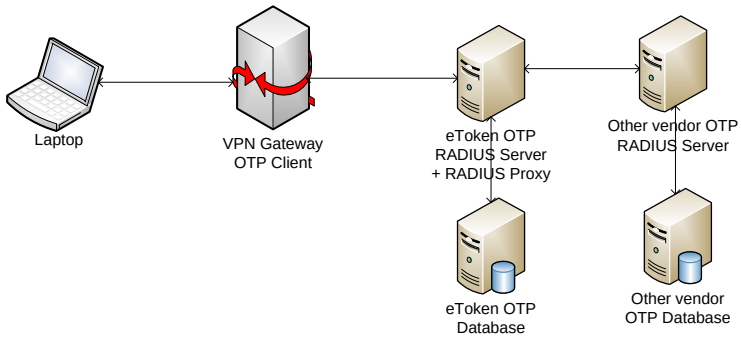
For organizations already using OTP provided by other vendors, it is possible to migrate the users to eToken OTP.

The following figure illustrates an OTP client sends an authentication request to the other vendor's RADIUS server.

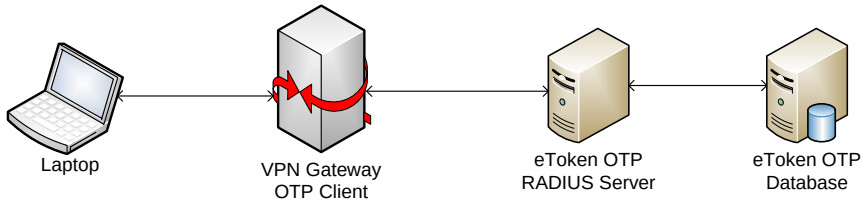


As other vendors' tokens are being used in the organization, a migration process must be implemented where both tokens are used during the migration phase, until the migration process is complete.

The following figure illustrates the OTP client sending all requests to the eToken OTP RADIUS server, which first answers all of its own token OTP requests, and then proxies all the other requests to the other vendor's OTP RADIUS servers.



After all the tokens of the other vendor have been replaced by eToken OTP tokens, the other vendor's OTP RADIUS server can be removed from the network as illustrated in the following figure:



For an example of a migration process, see the eToken Integration Guide: *Migrating from RSA SecurID to eToken OTP Authentication*.

TMS Remote Site for Internet/Extranet

If the *TMSRemote* web site is to be used, it is recommended to put the server on the DMZ network.

To secure the server, it is recommended to strengthen the web site security and set it to SSL, in addition to removing all other TMS web sites from the IIS except *TMSRemote*. The *TMSManage*, *TMSService*, *OTPAuthentication*, *TPOManagement* and *TMSAgent* virtual directories must all be removed.

When end-users who are on are not on location access the *TMSRemote* web site, will not enter network credentials to log in, but are required to answer predefined security questions and submit a CAPTCHA password to reduce brute force attacks.

For more information on the *TMSRemote* web site, see *eToken TMS 5.1 Administrators Guide*.

Backup and Restore

After installing the TMS server, it is important to backup the TMS contents so that it is possible to restore in the event of any failure.

The backup needs to include:

- The actual TMS Database, i.e. ADAM.
- The security keys.
- The roles configuration.

For information on how to backup and restore the TMS database, see KB article 1487, *TMS Backup and Restore*.

For information on how to backup and restore the TMS security keys and roles configuration, see *TMS 5.1 Administrators Guide*.

Selecting the Integration Technology

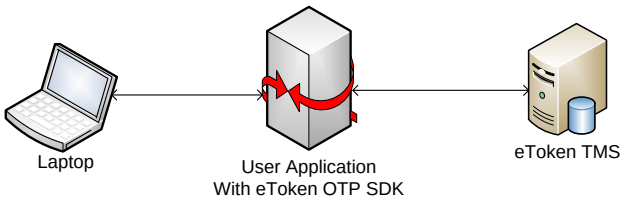
Most OTP clients on the market today support the RADIUS protocol for authentication, which automatically enables application enhancement to use OTP strong authentication.

Where a desktop or web application does not have RADIUS support there are two possible options to add OTP support:

- Using the eToken OTP Software Developer Kit (SDK)
- Adding RADIUS support to the application

The eToken OTP SDK enables the addition to the application a component that sends an authentication request for user validation directly to the TMS Server, without contacting a RADIUS server.

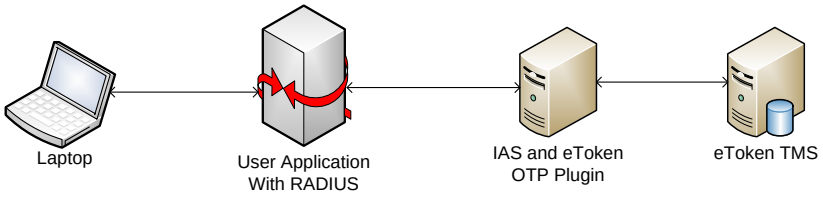
The advantage of using the eToken OTP SDK is that it is easy to implement, as illustrated in the following figure:



Alternatively, adding RADIUS support to the application provides more flexibility, since it gives the ability to use any RADIUS server.

Another advantage when using a RADIUS server is that the RADIUS server has more flexibility with configured policies, such as answering requests within a certain time or with certain attributes.

The following figure illustrates the RADIUS authentication procedure:



Life Cycle Management Using the TMS OTP API SDK

The Help Desk team and end-users can manage tokens using the TMS portals.

Alternatively, application developers can integrate some token lifecycle management operations into existing or new web applications, using the eToken TMS OTP API SDK.

The eToken TMS OTP API provides the following functionality:

- Authenticate a user with OTP
- Unlock an OTP token
- Change or reset the OTP PIN

- Use a temporary password
- Validate OTP
- Assign an OTP token

The eToken TMS OTP API SDK package can be obtained from your local distributor.

Understanding Redundancy and Scalability

This section describes the issues to consider when implementing eToken TMS in a large organization where both scalability and redundancy are required.

Total Number of Users and Concurrent Requests

When planning the architecture of an eToken TMS installation, it is important to distinguish between the requirement to support large user stores and the requirement to support a large number of requests in a short time period. Supporting an organization with, for example, one million users, but with a very low request ratio, requires a totally different architecture from an organization with a thousand users who are frequent users.

Supporting a large user store usually requires little more than large disk space. With large user stores it is important to avoid operations such as enumeration and report generation which return large sets of user data. For example, enumerating one million users takes a long time and can cause the application to stop responding.

Fortunately, such operations are not usually required as both management and authentication processes are performed for a specific user or token.

By contrast, support for many concurrent users (or requests) does require strong computation power and additional architectural components.

Failover Options in eToken TMS

Failover is the capability to switch over automatically to a redundant or standby computer server, system, or network upon the failure or abnormal termination of the previously active server, system, or network. Failover happens without human intervention and usually without warning. System designers usually provide failover capability in servers, systems or networks requiring continuous availability and a high degree of reliability.

As eToken TMS functionality is provided in the form of web applications (either web sites or web services), failover is supported by the well-known model of web server failover. In this model, more than one eToken TMS server is installed in the organization and if one of the servers fails to operate, another server activates and fulfills the requests.

Standard web application tools allow all servers to have a single name, so that the end-user (administrator or application in case of a web service) can always access the same web address while actually being served by different servers based on server availability.

This achieves eToken TMS application availability, but eToken TMS also relies on user and configuration stores to function. For this, eToken TMS relies on AD and ADAM replication capabilities provided by Microsoft, which support data storage in multiple computers. Data is synchronized between all computers and any read, query or write operations are performed using the most responsive server available, determined by server load and network availability.

eToken TMS OTP 2.5 and later supports redundancy by enabling connection to several RADIUS servers. If one RADIUS server fails, another one can take over.

Load Balancing Options in eToken TMS

Load balancing is a technique that distributes processing and communications activity across a computer network so that no single device is overwhelmed. The use of multiple components with load balancing, instead of a single component, increases reliability through redundancy. The balancing service is usually provided by a dedicated program or hardware device.

Load balancing is especially important for networks where it is difficult to predict the number of requests a server is required to process.

In eToken TMS, load balancing is required when a single server cannot process the number of expected requests at peak times. This is usually due to authentication, as the number of users requiring authentication to the organization's gateways is usually much larger than the number of management operations performed at a particular time.

A single eToken TMS server is usually enough to support the number of requests in a small, medium and even large organization (with up to tens of thousands of users). However, in a high end environment, load balancing is required.

eToken TMS relies on web load balancing. Standard RADIUS load balancing tools and applications (such as Alteon) can be used to support an eToken TMS deployment configuration.

While failover is required by almost every organization, it is recommended to implement load balancing. Most small and medium organizations do not require load balancing.

Calculating Expected Authentication Requests

The load created by authentication requests is dependent on the following factors:

- The number of users with a token.
- The number of token users who need to log on from outside the company premises. The usage can be greater than when the same users are working in the office.
- Peak periods when the number of token authentications is high. For example, at the start of the working day or in the evening after the end of the standard working day.

It is recommended to implement a small system and generating RADIUS reports to show the number and time of token authentications. The system can then be extended if required.

Understanding Network and User Store Latency

A single request response time (the length of time between the generation of the request and the return of the response) is often considered to be dependent on the server availability and load. A free server answers quickly while busy and a loaded servers require more time. In complex network environments the time required for actual computation and processing time on the server is less significant in the overall response time, than the time spent on network round-trips from the client to the application server and to the different stores.

When trying to resolve a slow eToken TMS server response time, the first item that that should be evaluated is the network latency and network resources (such as stores). For example, if a network round-trip of client-server-store requires one second, no authentication request can occur faster even if the eToken TMS can compute the result in near zero time.

In addition, there is no linear connection between the time it takes for a single request to be processed and the number of requests that can be handled in a specific period of time. This is due to parallel processing of all components involved including NAS applications, eToken TMS server and the different stores.

Parallel processing can be illustrated by the analogy of a fast food restaurant: while a single person waits five minutes for his burger, many customers can be served in the same time period by having many service stations.

When analyzing system performance the following should be verified:

- A single request response time should be no higher than the time a typical user would be willing to wait which is usually one to two seconds. If network latency is low, this time depends only on eToken TMS computation and this is usually achieved. If network latency is high, then any investment in upgrading eToken TMS server computation capabilities will be ineffective.
- The overall number of requests that the system can process in a period of time should be higher than the expected processing requirements at peak times. This parameter is almost not affected by network latency and depends on eToken TMS and system computation capability. As previously explained, even a single eToken TMS server can serve the needs of an organization with tens of thousands of users.

High end environments should consider using load balancing features.

Achieving Scale and Redundancy in Multi-site Environments

Multi-site environments, common in enterprise organizations, often include both very fast networks (within each site) and networks with very high latency (the links between sites). Such an environment has both the following redundancy and load balancing characteristics:

- As the link between two different sites is not guaranteed, to operate continuously, it is usually required to keep a local eToken TMS service on each site, at least for circumstances where a central service is not available.
- As the number of sites usually implies a large number of users, such an environment requires load balancing features.

When designing such an environment, the following should be considered:

- If the link between two sites is considered stable and is used by other high availability network applications, then eToken TMS can safely reside on only one of the sites because eToken TMS provides web services which pass only small amounts of data (unlike thick client applications).
- If the link between sites cannot be guaranteed to be operational at all times, a local eToken TMS service should be available on each site. This service itself can have failover and load balancing capability.

Recommended Architecture Options

In the single-server all-in-one configuration, all the system components including user stores and eToken TMS stores, eToken TMS server, and the IAS server are installed on a single computer. To support a distributed, multi-site environment, a distributed solution with high availability and load balancing must be considered. This section describes the recommended architecture and component layouts.

Hardware Requirements

The following table describes the hardware requirements for a single TMS server installation:

Component	Minimum	Recommended
Processor	2.5 GHz	Dual processors, each processor 3 GHz or faster
RAM	2 GB	More than 2 GB
Disk	NTFS-formatted partition with a minimum of 3 GB of free space	NTFS-formatted partition with 3 GB of free space, plus adequate free space for data storage
Drive	DVD-ROM drive, or a local or network-accessible drive to which entire contents of the TMS installation setup can be copied	

Component	Minimum	Recommended
Display Monitor	Resolution: 1024 X 768	Resolution: 1024 X 768 or higher
Network Bandwidth (for network computer connections)	100 Mbps	1 Gbps

Software Requirements

The following table describes the software requirements for a TMS Server:

Component	Requirement	Comments
Supported Operating Systems	Windows Server 2003 with SP2 (32-bit or 64-bit)	
Additional Software	Windows Installer 3.0 or higher	The Microsoft Windows Installer is an application installation and configuration service. WindowsInstaller-*KB884016-v2-x86.exe* is the redistributable package for installing or upgrading Windows Installer. http://www.microsoft.com/downloads/details.aspx?familyid=5fbc5470-b259-4733-a914-a956122e08e8&displaylang=en
	Microsoft .NET Framework Version 2.0 SP1 Redistributable	The Microsoft .NET Framework version 2.0 SP1 (x86) redistributable package installs the .NET Framework runtime and associated files required to run applications developed to target the .NET Framework v2.0 SP1. http://www.microsoft.com/downloads/details.aspx?familyid=0856EACB-4362- 4B0D-8EDDAAB15C5E04F5&displaylang=en
	One of the following: ♦ Microsoft SQL Server 2005 ♦ Microsoft SQL Server	Required if the attendance report function is used.

Component	Requirement	Comments
	2008	
	Java Runtime Environment 1.5 or higher Cabarc.exe (Microsoft Cabinet Tool)	Required if eToken MobilePass is used.
TMS User Store	TMS requires one of the following as TMS External User Store: ♦ Active Directory (Windows 2003 or 2008) ♦ MS SQL Server 2005 or 2008 ♦ OpenLDAP 2.3.38 or higher ♦ Novell eDirectory 8.7.3 or higher Or Internal User Store ♦ TMS ADAM Internal	MS SQL Server does not support the Badging feature. **Note:** If the integrated configuration and user store is used, ADAM is installed during the TMS installation and a configuration store does not need to be pre-installed.
eToken Applications	eToken PKI Client, version 4.55 or later. Required only for eToken NG-OTP enrolment.	

Basic Deployment

This section describes the following architecture scenarios:

- Single server all-in-one
- Distributed: TMS with separate directory server cluster
- Distributed: TMS with separate RADIUS and directory server cluster

Single Server All-in-One

Architectural Layout

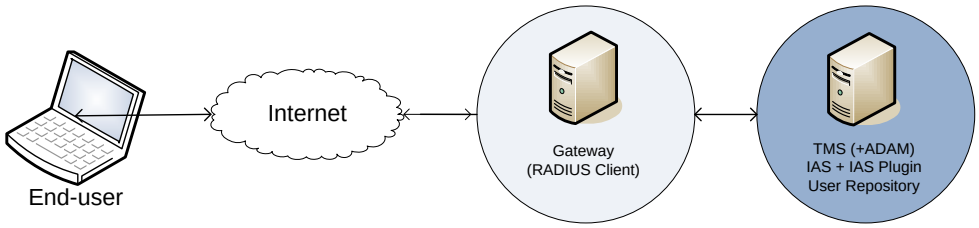
In this scenario, eToken TMS is installed in the same server as the RADIUS server and the user repository store services.

Installing all the system components on the same server is the simplest and most convenient configuration and meets the needs of most organizations. The proximity of all system components minimizes network latency and enables easy entire system management.

For small organizations that do not require 24 hour availability, this is the recommended scenario.

Note: Using AD as a cluster is the most widely used configuration, because the AD cluster is often in place before the deployment of eToken TMS.

Solution Diagram



Software Requirements

In this scenario, one server is installed with the following:

- eToken TMS Server
- Microsoft ADAM (optional)
- Microsoft IAS with Aladdin eToken IAS plug-in
- One of the following:
  - Microsoft Active Directory
  - Microsoft SQL Server
  - OpenLDAP
  - Novell eDirectory
  - TMS Internal user repository

Distributed: TMS with Separate Directory Server Cluster

Architectural Layout

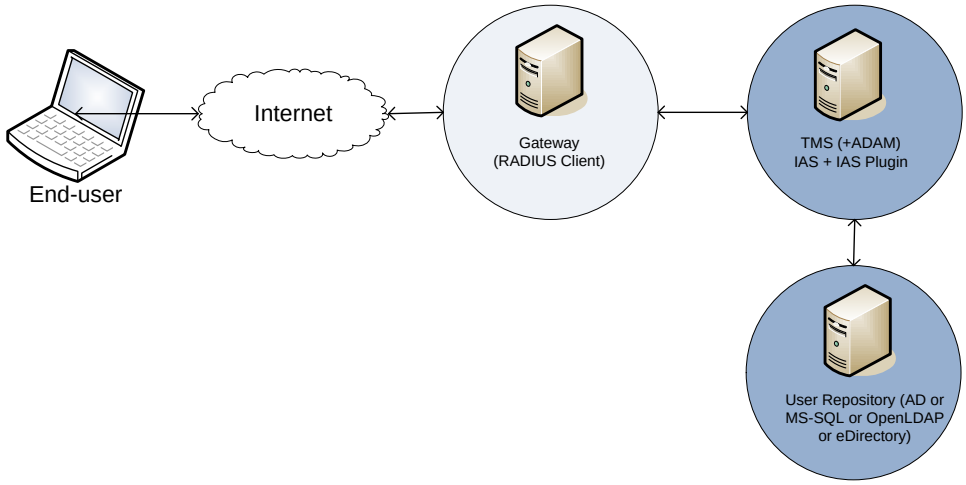
As eToken TMS is installed in an environment where a user store already exists, usually running on a cluster of directory servers (Domain Controllers in a Microsoft environment), a common variation of the All-In-One model is to install all components except the user store, on a single machine and connect this system to the existing user store cluster.

This is also implemented when MS SQL Server, OpenLDAP or eDirectory or TMS internal user repository is used as the user store; connect the all-in-one machine to the existing user store.

- *eToken TMS rule of thumb #1* – When installing eToken TMS in a new clean environment, use the All-In-One method and install all eToken TMS components on a single machine.
- *eToken TMS rule of thumb #2:* If a user store already exists, connect to it from the installed machine; or install the store on the same machine.

For small organizations that already have a user repository cluster in a different sever or servers and do not require 24 hour availability, this is the recommended behavior.

Solution Diagram



Software Requirements

In this deployment, the TMS server is installed on one machine, and the user repository server exists on other server or servers.

- TMS:
  - eToken TMS Server + Optional ADAM
  - Microsoft IAS + eToken IAS plug-in
- User repository: On a separate server or a cluster of servers, one of the following is used:
  - Microsoft Active Directory
  - Microsoft SQL Server
  - OpenLDAP
  - Novell eDirectory
  - TMS Internal user repository

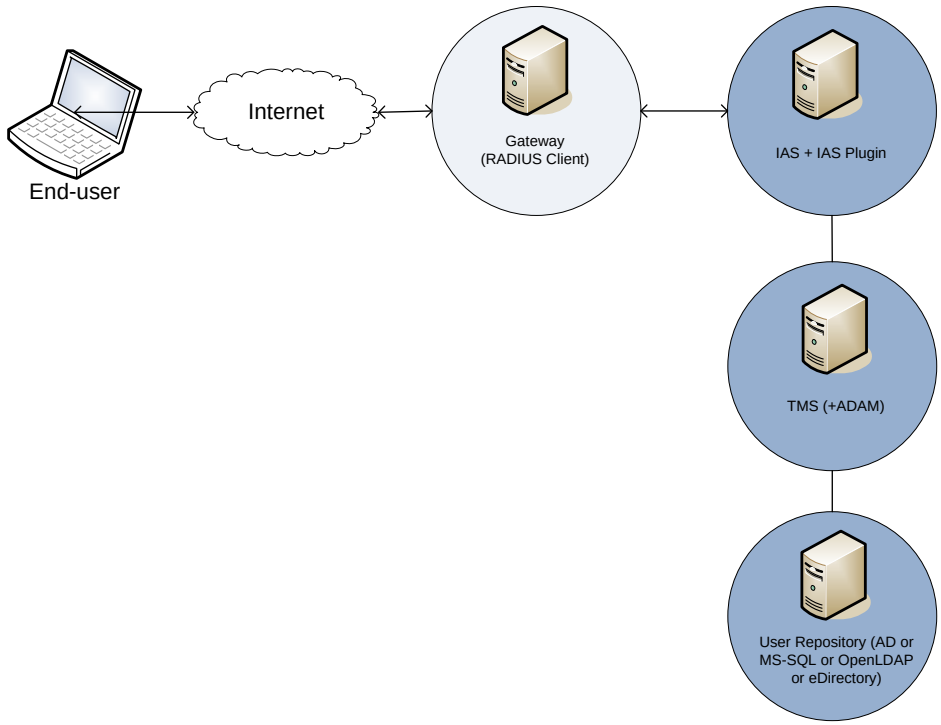
Distributed: TMS with Separate RADIUS and Directory Server Cluster

Architectural Layout

In this scenario, there is an additional step in the separation of the services. One server is used as a RADIUS server, the second as the TMS server and the third as the user repository (single server or cluster of servers).

This is the recommended configuration for an organization that already has an IAS server being used for different authentication services and cannot be moved.

Solution Diagram



Software Requirements

- RADIUS Server:
  - Microsoft IAS + eToken IAS plug-in
- TMS Server:
  - eToken TMS Server 5.1
  - Microsoft ADAM (optional)
- User repository - one of the following:
  - Microsoft Active Directory
  - Microsoft SQL Server
  - OpenLDAP
  - Novell eDirectory
  - TMS Internal user repository

Load Balancing and Fault Tolerance

Load balancing is a technique to spread work between two or more servers to get optimal resource utilization, maximize throughput, and minimize response time, in addition to fault tolerance which may caused by a server failure. Using multiple components with load balancing, instead of a single component, can increase reliability through redundancy. The balancing service is usually provided by dedicated software, for example, Microsoft Network Load Balancing (NLB) or dedicated hardware, such as an NLB appliance.

For TMS services, the load balancer is usually a software program which listens on the port where external clients connect to access services. The different ports are as follows:

- RADIUS port – If the NLB is before the RADIUS
- HTTP or HTTPS ports – If the NLB is before the TMS server (since TMS services are web services which work with HTTP or HTTPS).

The load balancer then forwards the requests to one of the "backend" servers, which usually replies to the load balancer. This enables the load balancer to reply to the client without the client ever knowing about the internal separation of functions. It also prevents clients from contacting backend servers directly, which can have security benefits by hiding the structure of the internal network and preventing attacks on the kernel's network stack or unrelated services running on other ports.

Most load balancers have the ability to off-load SSL using an application-specific integrated circuit. By moving the SSL negotiation to the front side of the appliance the SSL workload can be distributed and the traffic examined, providing an additional application security layer.

Some load balancers provide a mechanism for doing something additional in the event that all backend servers are unavailable. This can include forwarding to a backup load balancer, or displaying a message indicating the outage.

Duplicating the TMS Server

To duplicate a TMS server to another machine, perform the following:

1. Install a new eToken TMS server.
2. Export the security keys from the original eToken TMS server to the duplicated eToken TMS server.
3. Point all eToken TMS servers to the same Roles database.

Deployment with Redundancy (Single Site)

In the following scenarios, using the All-in-One building block is still recommended, but with these building blocks being distributed differently. The following configurations include an All-in-One cluster on the same site with an optional separated user repository server or cluster of servers.

The replication of the stores' data relies on the replication capabilities of both user store and eToken TMS configuration store directories (either AD or ADAM).

Single Server All-in-One with NLB

Architectural Layout

In this scenario, eToken TMS is installed in the same server as the RADIUS server and the user repository store services.

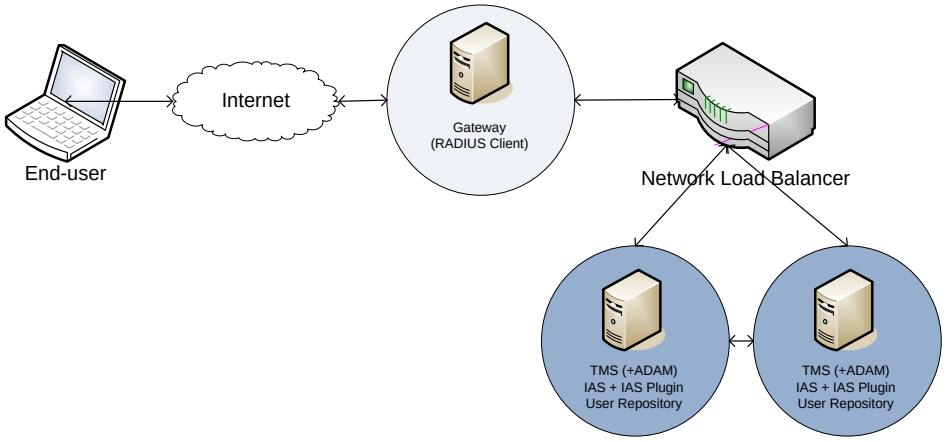
Installing all the system components on the same server is the simplest and most convenient configuration and meets the needs of most organizations. The proximity of all system components minimizes network latency and enables easy entire system management.

To create load balancing for performance and fault tolerance for availability, use multiple TMS Servers working together in a Network Load Balancing (NLB) environment.

This is the recommended scenario for organizations that need high availability for the OTP usage in the organization.

Solution Diagram

The multiple IAS/TMS servers are connected by a network load balancer (NLB).



Software Requirements

Each server includes the following components:

- eToken TMS Server
- ADAM (optional)
- User repository - one of the following:
  - Microsoft Active Directory
  - Microsoft SQL Server
  - OpenLDAP
  - Novell eDirectory
  - TMS Internal user repository

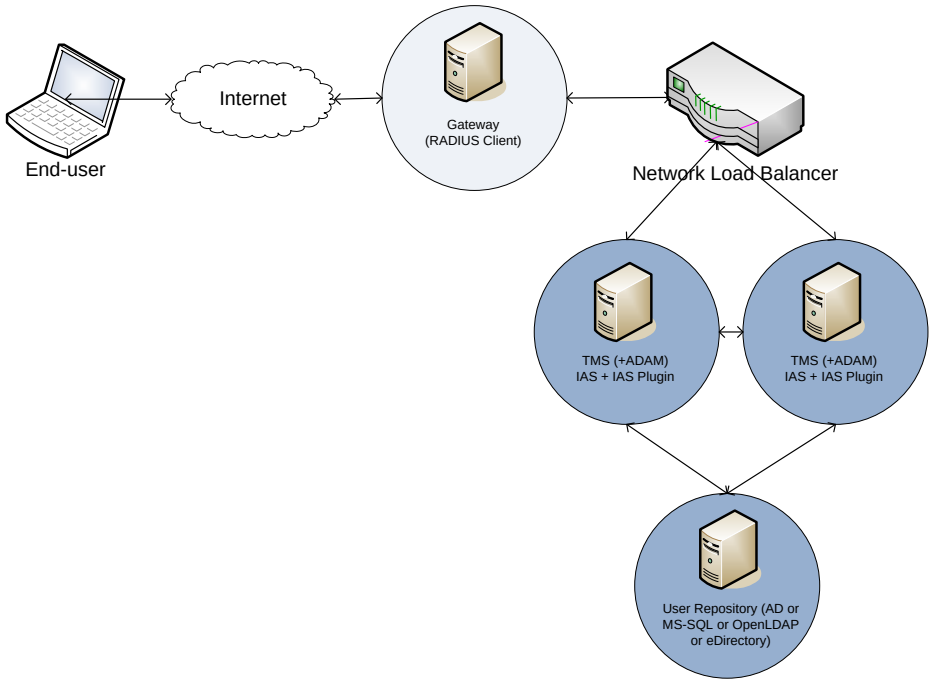
Distributed: TMS with Separate Directory Server Cluster with NLB

Architectural Layout

In this scenario, eToken TMS is installed in the same server as the RADIUS server with a separate directory server or servers cluster, with NLB for load balancing and fault tolerance.

This is the recommended scenario for organizations that need high availability for the OTP usage, when the organization already has a separated directory server or servers cluster that already exists.

Solution Diagram



Software Requirements

- Each server includes the following components:
  - eToken TMS Server
  - Microsoft ADAM (optional)
- User repository server/servers - one of the following:
  - Microsoft Active Directory
  - MS SQL Server
  - OpenLDAP
  - Novell eDirectory
  - TMS Internal user repository

Distributed: TMS with Separate RADIUS and Directory Server Cluster with RADIUS Redundancy

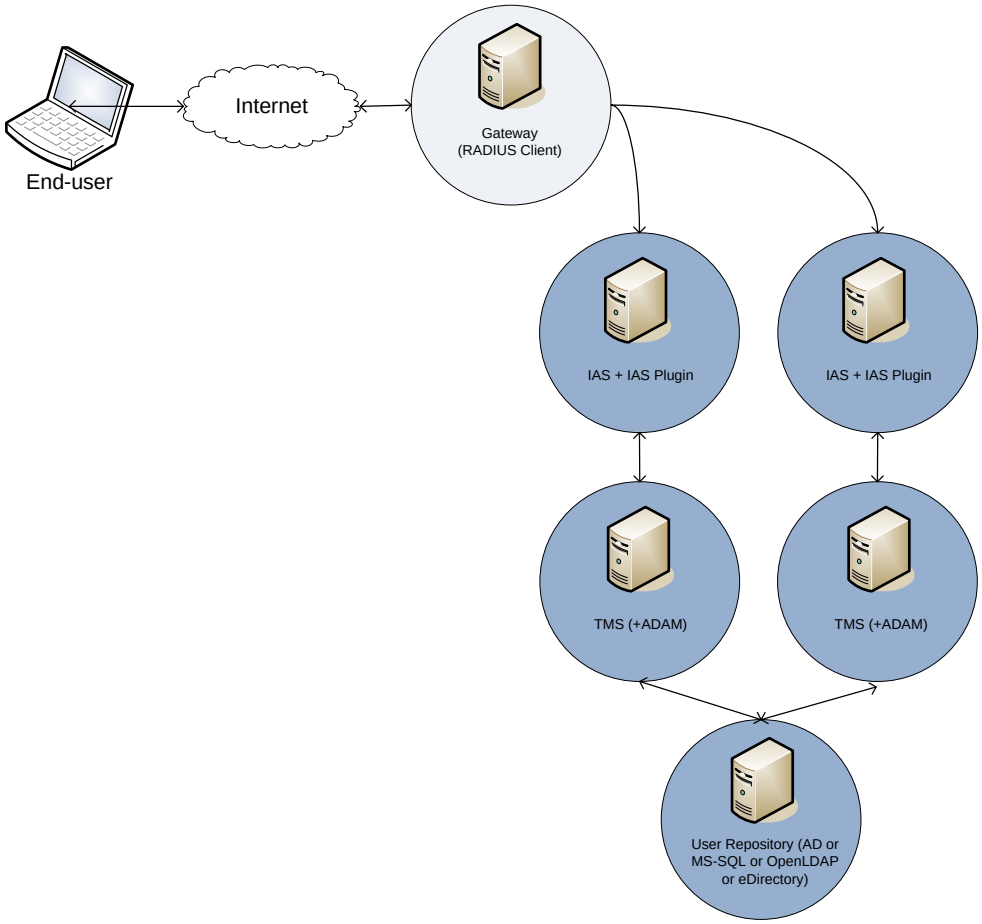
Architectural Layout

In this scenario RADIUS redundancy is used instead of NLB.

Most gateways that have the RADIUS client module for authentication can configure more than one RADIUS server for authentication. This can be used for load balancing and fault tolerance.

In the gateway two or more RADIUS servers are configured, and in the event that the first RADIUS server does not respond, the gateway sends an authentication request to the second RADIUS server, and so on.

Solution Diagram



Software Requirements

- RADIUS Server:
  - Microsoft IAS + eToken IAS plug-in
- TMS Server:
  - eToken TMS Server 5.1
  - Microsoft ADAM (optional)

- User repository - one of the following:
  - Microsoft Active Directory
  - MS SQL Server
  - OpenLDAP
  - Novell eDirectory
  - TMS Internal user repository

Deployment with Redundancy in a Multi-site Environment

This section describes the following scenarios:

- Single Server All-in-One with NLB in a multi-site environment
- Distributed: TMS with separate directory server cluster with NLB in a multi-site environment

Single Server All-in-One with NLB in a Multi-site Environment

Architectural Layout

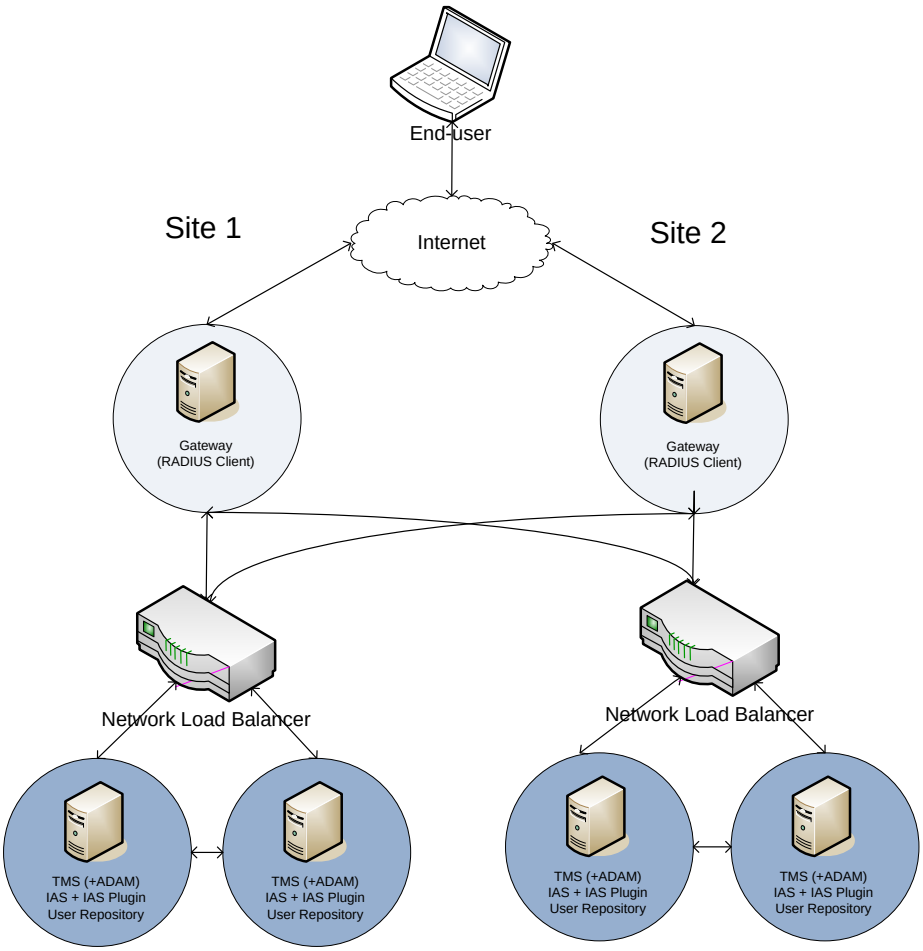
In this deployment, the *Single Server All-in-One with NLB* scenario is repeated on each site in the enterprise. Each site is connected to other sites, as determined by the enterprise's network topography, enabling redundancy and load balancing between sites.

This multi-site dependency is recommended for extreme failover cases. The users of each site use the closest cluster and are redirected to a remote one when the entire local cluster is not operational.

In this configuration low network latency can be achieved by having each client connect to the closest cluster while serving many clients over the entire organization.

If the system is distributed, replicate the All-In-One cluster for as many sites as required. Keep a remote site in the failover list as a last resort, in case the entire local cluster is not operational.

Solution Diagram



Software Requirements

Each server includes the following components:

- eToken TMS Server
- Microsoft ADAM (optional)

- User repository - one of the following:
  - Microsoft Active Directory
  - Microsoft SQL Server
  - OpenLDAP
  - Novell eDirectory
  - TMS Internal user repository

Distributed: TMS with Separate Directory Server Cluster with NLB in a Multi-site Environment

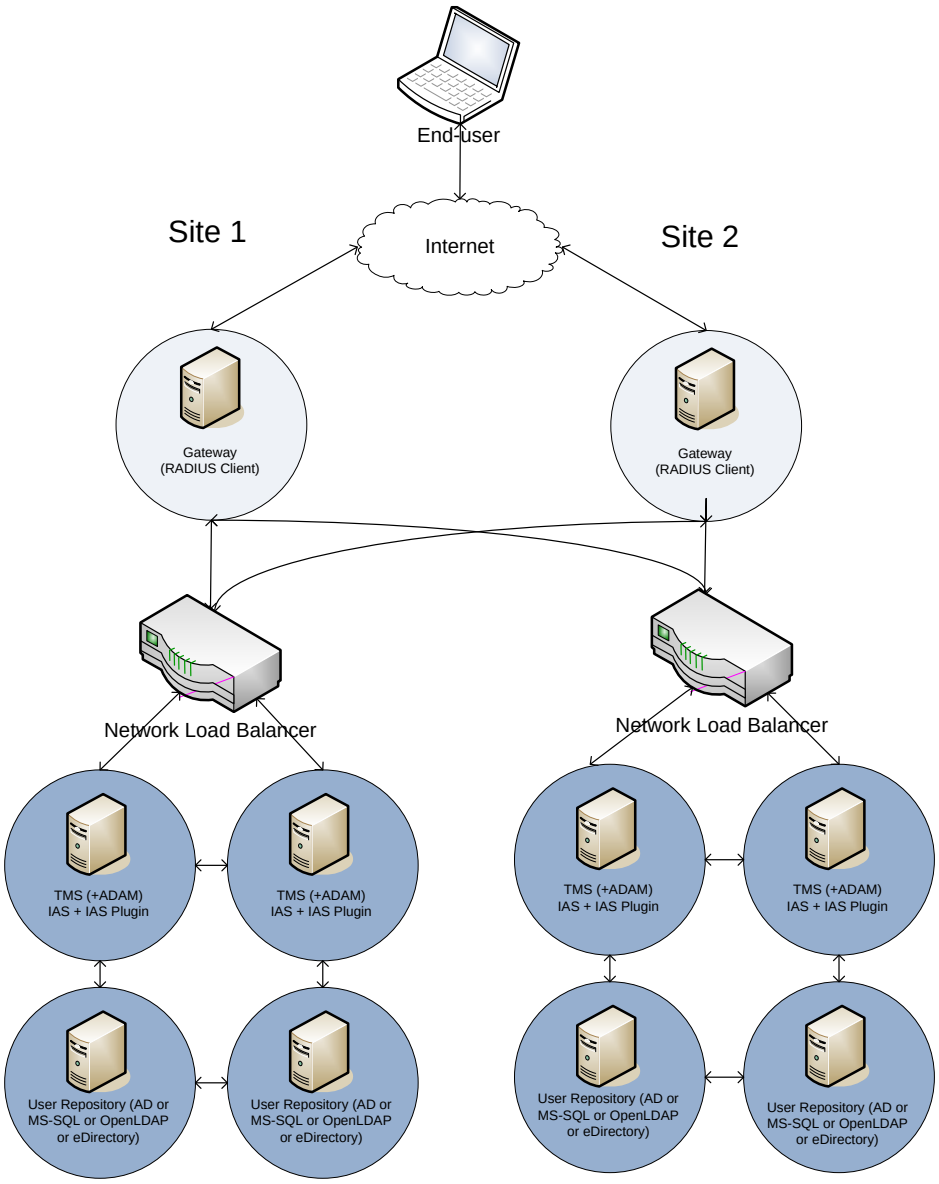
Architectural Layout

In this deployment, the *Single Server All-in-One with NLB in Multi Site* is repeated, but with a separated directory server cluster. The load balancing is repeated on each site in the enterprise. Each site is connected to other sites, as determined by the enterprise's network topography, enabling redundancy and load balancing between sites.

This multi-site dependency is recommended for extreme failover cases. The users of each site will use the closest cluster and be redirected to a remote one when the entire local cluster is not operational.

This is the required option if the user repository database is located on different servers.

Solution Diagram



Software Requirements

- Each server includes the following components:
  - eToken TMS Server
  - ADAM (optional)

- A user repository exists on a different cluster:
  - Microsoft Active Directory (AD)
  - Microsoft SQL Server
  - OpenLDAP
  - Novell eDirectory
  - TMS Internal user repository

Chapter 5

Sizing and Performance

This chapter contains the highlights of performance measurements for eToken TMS OTP authentication as performed in SafeNet's Authentication Performance Lab. It also provides advice on ways to increase OTP authentication performance.

In this chapter:

Capacity Planning Considerations

The measurements described in this section are to be used as a planning guidance for the implementation of OTP authentication solutions. Each deployment of eToken TMS must be assessed on a case by case basis.

Note: SafeNet provides no guarantee that the same performance can be achieved in each and every case, as actual sustained authentication performance in an organization is affected by many factors, such as network speed, latency, and load caused by other network traffic.

Performance Lab Testing Methodology

The performance matrix as described in the following paragraph Performance Matrix by Configuration includes performance measurements conducted using different hardware configurations. In all the following cases the systems used for the test is as follows:

- Microsoft Windows 2003 Server SP2
- 64-bit platform
- Running eToken TMS 5.1
- User and authenticator repositories tested were Active Directory and ADAM respectively.

In some of the tests, Network Load Balancing equipment was used to increase performance.

Testing Lab Configuration Specification

The following tests were conducted on the following hardware platforms:

- Standard Server - IBM X3250
  - Intel Xeon 3040 1.86GHz Dual Core
  - 4 GB RAM

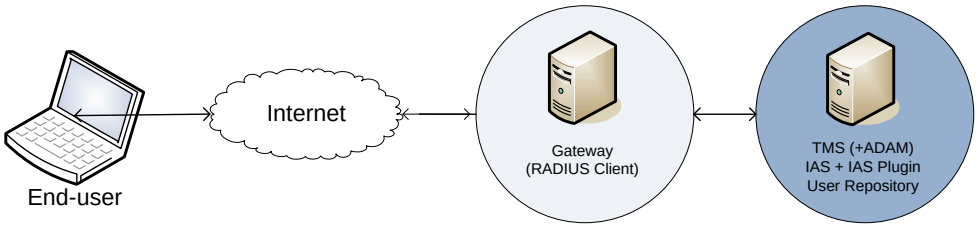
- Enterprise Server - HP ProLiant DL360 G5 Server series
  - Intel Xeon E5335 2.0GHz Quad Core
  - 4 GB RAM
  - HP 512MB w/Battery Smart Array Battery Back Write Cache Enabler
- Network Load Balancer - ALTEON 180e ACESwitch 8-PORT 1000 Base-SX WEB SWITCH

Performance Matrix by Configuration

The following measurements are based on RADIUS client authentication via a Microsoft IAS server configuration. The tests were based on an AD repository of 100,000 users and were conducted using 8 simultaneous RADIUS client sessions. Each client session tested 10,000 different users. Each complete single test was based on 80,000 users.

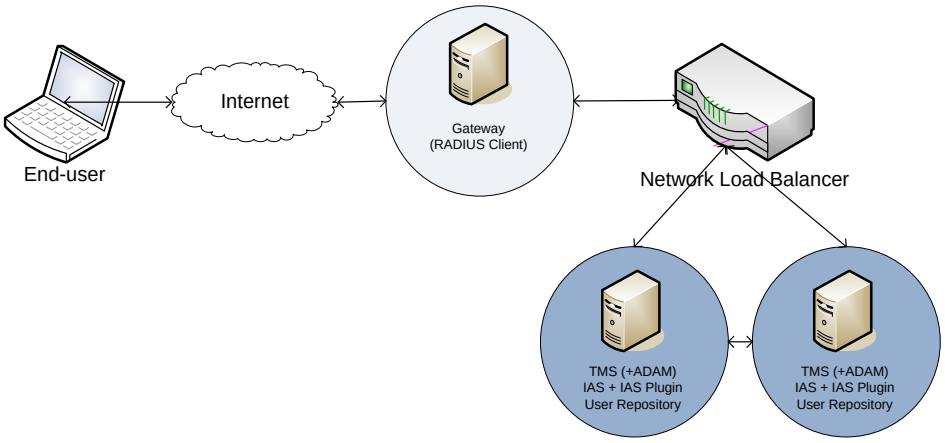
Note: It is anticipated that clients that authenticate using the eToken TMS OTP SDK, leveraging the OTP authentication web services, will provide even better authentication throughput.

Single Server All-In-One



	Auth per second (peak)	Auth per second (sus-tained)	Auth per hour (sus-tained)	Success auth (%)	Rejected auth (%)	Auth timeouts (count)	Auth latency
All in one	59	48	172800	100%	0%	0	20ms

Single Site Configuration All-in-One with NLB



	Auth per second (peak)	Auth per second (sus-tained)	Auth per hour (sus-tained)	Success auth (%)	Rejected auth (%)	Auth timeouts (count)	Auth latency
2 All-in-one +NLB	108	96	345,600	100%	0%	0	10 ms
3 All-in-one +NLB	156	138	496,800	100%	0%	0	7 ms
4 All-in-one +NLB	210	186	669,600	100%	0%	0	5 ms

Fine Tuning for Optimal Performance

Other methods to increase OTP authentication performance as much as possible is as follows:

- **TMS Configuration Tweaks:** For information on tweaking the performance of the TMS Server, refer to
  - *OTP Web Services Configuration* section in the *eToken TMS 5.1 Administrators Guide*.
  - OTP Performance Tuning section in the TMS 5.1 Troubleshooting Guide.
- **Hardware:**
  - Use the recommended hardware specifications as described in the *Hardware Requirements* section on page 51.
  - Improve disk I/O performance for the TMS server system by adding a Battery Backed Write Cache (BBWC) module to gain a performance boost.
- **Network:** Increase the latency of the authentication requests between the different servers by using fast 1GB network connections between them.
- **Architecture Considerations:** Enhance performance through eliminating network latency by putting as many services as possible on the same machine: RADIUS, TMS and user repository.

Appendix A

References

For more information, refer to the following guides or packages:

For more information on...	Refer to...
TMS 5.1 Installations and Configuration	eToken TMS 5.1 Administrators Guide eToken TMS 5.1 Users Guide eToken TMS 5.1 README
OTP IAS Plug-in and OTP Agents (Citrix, OWA, IIS)	OTP Authentication 5.1 Administrators Guide eToken OTP 5.1 README
Microsoft Integration	eToken Microsoft Integration Guide 2. 0
OTP SDK	OTP SDK 5.1 Guide
Tweaking TMS Performance	TMS 5.1 Troubleshooting Guide
SafeWord Migration	SafeWord migration to TMS 5.1 How to guide
TMS Backup and Restore	KB article 1487 TMS Backup and Restore