Detection And Hunting Of Golden SAML Attack

By Sygnia

Published: 2021-07-21 · Archived: 2026-04-05 21:54:43 UTC

The SolarWinds software supply chain attack is known to have affected U.S. government agencies, critical infrastructure entities, and private sector organizations by an advanced persistent threat actor since at least March 2020. U.S. authorities now believe that additional initial access vectors other than the SolarWinds platform exist, but these are still being investigated. The US Cybersecurity & Information Security Agency (CISA) expects that removing this threat actor from compromised environments will be highly complex and challenging.

One of the major techniques used by the threat actor as part of the SolarWinds attack, was compromising the Security Assertion Markup Language (SAML) signing certificate, using their Active Directory privileges. CISA explained that “once this is accomplished, the adversary creates unauthorized but valid tokens and presents them to services that trust SAML tokens from the environment. These tokens can then be used to access resources in hosted environments, such as email, for data exfiltration via authorized application programming interfaces (APIs)”[1].

The “Golden SAML” attack technique enables attackers to forge SAML responses and bypass ADFS authentication to access federated services. First reported by CyberArk in 2017, the current attack is the first time that this technique is known to have been used “in the wild”.

To successfully leverage Golden SAML, an attacker must first gain administrative access to the ADFS server and extract the necessary certificate and private key. Once this is accomplished, unauthorized access can be performed from anywhere, without further access to the victim environment.

Unless discovered and remediated, this attack provides attackers with persistent access to all services federated with ADFS. Such services often include critical infrastructure and sensitive data such as AWS and Office 365. Accounting for some of the key functions and systems that commonly use SAML, CISA referred to hosted email services, hosted business intelligence applications, travel systems, timecard systems, and file storage services (such as SharePoint).

The recent rise in awareness and novel in-the-wild use of this attack significantly raises the likelihood of attackers leveraging it to their advantage. It is therefore highly advised that organizations move swiftly in taking the necessary steps to protect their SSO infrastructure and establish effective monitoring to detect and respond to such attacks.

Golden SAML Attack

In order to explain the Golden SAML attack, we’ll first provide a short description of the legitimate SAML authentication process. The process involves the following steps, illustrated in **Figure 1**:

1. User attempts to access desired service (e.g. AWS, Office 365).

2. Service redirects user to ADFS for authentication.
3. User authenticates with ADFS according to Domain policy (e.g. Multi-Factor-Authentication).
4. ADFS returns signed SAML response to user machine.
5. User presents desired service with signed SAML response and receives access.

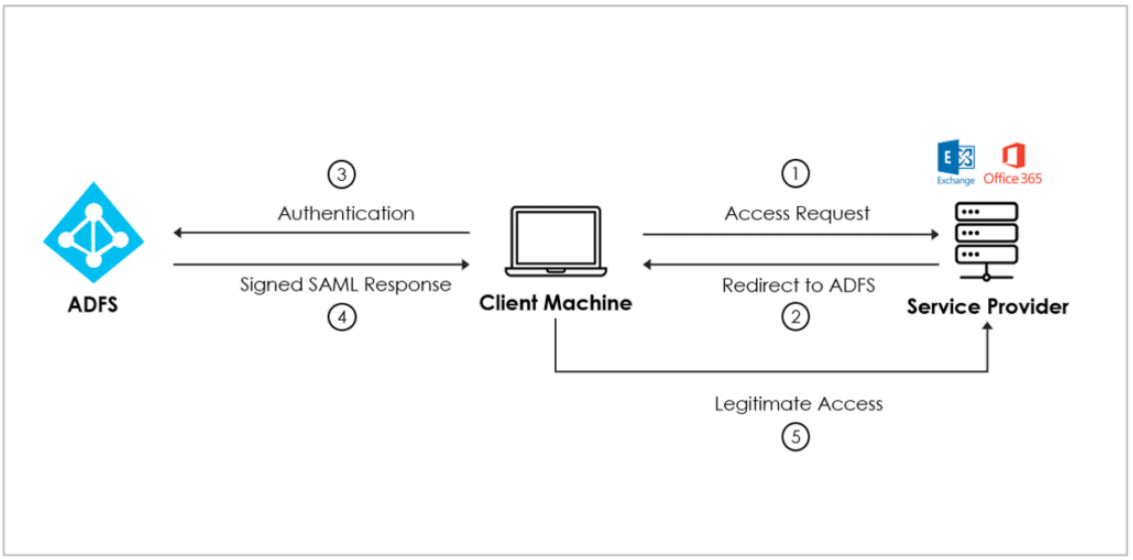

Figure 1: Legitimate SAML Authentication Process

When performing a golden SAML attack, an adversary must first gain administrative privileges on the ADFS server through additional Lateral Movement and Privilege Escalation. Once these privileges are obtained, the attack will proceed according to the following steps, illustrated in **Figure 2**.

1. Attacker accesses ADFS server and extracts private key and certificate.
2. User attempts to access desired service (e.g. AWS, Office 365).
3. Service redirects attacker to ADFS for authentication.
4. Bypassing ADFS authentication, attacker signs a forged SAML response with stolen key.
5. Attacker presents desired service with signed SAML response and receives access.

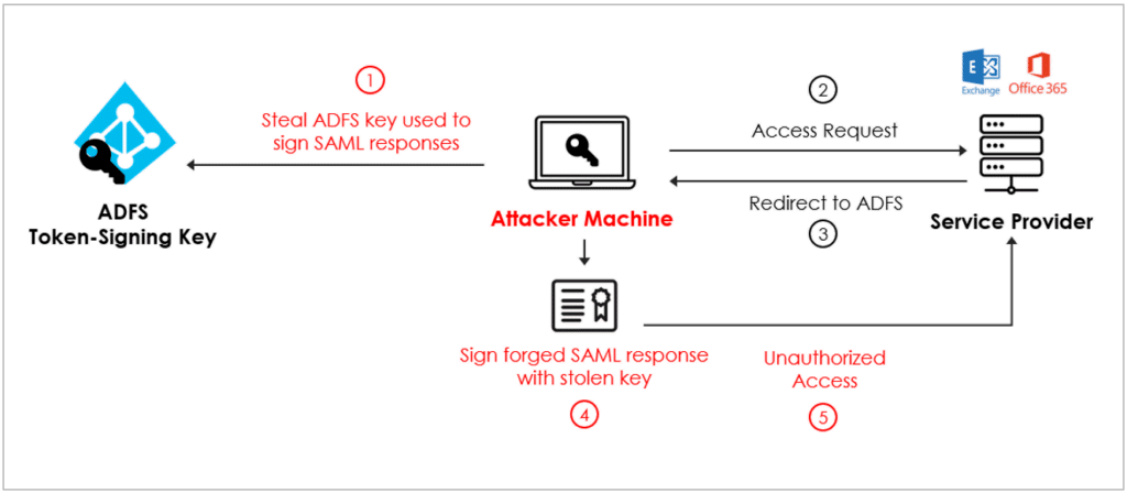

Figure 2: Golden SAML Attack Process

As can be seen in the attack process, once the ADFS private key and certificate are compromised attackers will gain persistent access to all relevant services. Critical data and infrastructure can now be compromised without any additional access to the victim environment. This powerful access will persist until the existing ADFS private key is invalidated and replaced, a difficult process requiring altering or terminating connectivity to all federated systems and re-creating them.

The increasingly rapid transition of critical assets to cloud services is making ADFS SAML and other SSO infrastructures lucrative high value targets for threat actors. The current hype and increased attention may lead to a rise of Golden SAML attacks. Along with the difficulty of remediating such attacks and the persistent access afforded to a successful attacker, we expect this technique to present a major challenge for defenders, SOCs and security teams.

Detecting and hunting Golden SAML attack

Based on understanding the attack process and modeling the attack in several environments, we recommend the following analyses to detect a Golden SAML attack. While there are many potential detection mechanisms which are only applicable when attackers erroneously leave unnecessary traces, our focus in this section will be on analyses capable of detecting any standard Golden SAML attack.

The analyses proposed here are designed for a standard Golden SAML attack, targeting an on-prem ADFS architecture. Other variations of SAML attacks, such as those targeting Azure AD, may require additional detection analyses.

Detection Method 1 – Correlating service provider login events with corresponding authentication events in ADFS and Domain Controllers

When performing legitimate authentication to an on-prem ADFS, the following events are logged:

In the ADFS server Windows Security Event Log:

- Event id 1202 – “The Federation Service validated a new credential”.
- Event id 1200 – “The Federation Service issued a valid token”.

In the Domain Controller Windows Security Event Log:

- Event id 4769 – "A Kerberos service ticket was requested", the source IP being the ADFS.

In the service provider logs:

- Appropriate log for each service, indicating a successful login (e.g. 'ConsoleLogin' or 'AssumeRoleWithSAML' in AWS or 'Sign-ins' in Azure AD).

However, when authenticating with a Golden SAML forged response the only event logged is the login to the service provider. The ADFS and Domain Controller are not involved in the authentication process in this scenario, and therefore do not log any events. While an attacker may also create a TGS request leaving behind a Kerberos requests (Event id 4769) on the Domain Controller, they cannot successfully authenticate to ADFS and create the standard authentication events.

Therefore, in order to detect Golden SAML authentications we can simply search for any logins to service providers using SAML SSO, which do not have corresponding 4769, 1200 and 1202 events in the Domain.

This detection mechanism is powerful, as it strikes at the core difference between a legitimate SAML authentication and a Golden SAML attack.

In order to facilitate this analysis, ensure the above mentioned events are enabled in audit policy:

- Audit ObjectAccess: success and failure on ADFS for event ids 1200 and 1202.
- AccountLogon – Kerberos Service Ticket Operations: success and failure on the Domain Controllers for event id 4769.
- Applicable login audit logs from service provider.

Detection Method 2 – Identifying certificate export events in ADFS

As previously mentioned, a Golden SAML attack requires the private key and certificate of the ADFS. Access to these can be obtained by exporting the certificate from the ADFS server, generating event id 1007 (enabled by default) in the 'Microsoft-Windows-CertificateServicesClient-Lifecycle-System' Windows Event Log.

Additionally, command-line evidence for exporting the ADFS certificate is likely incriminating. In many cases these can be found in:

- PowerShell script block logs: 'Export-PfxCertificate' or 'certutil -exportPFX' in event ids 4103 and 4104.
- Command line auditing tools (e.g. event id 4688): using certutil.exe -exportPFX.
- Execution evidence of known tools such as Mimikatz and ADFSDump. Certificate extraction with ADFSdump can be detected using Sysmon Event id 18, when the pipe name is "\microsoft##wid\tsql\query", excluding processes regularly making this pipe connection on the machine.

Detection Method 3 – Customizing SAML response to identify irregular access

In order to facilitate an additional layer of monitoring, organizations can modify their SAML responses to include custom elements for each service provider. Once enabled, these custom elements can be monitored in service

provider access logs to detect any anomalous requests. While these parameters can be faked by an attacker, they're likely to be different and facilitate effective detection of Golden SAML attacks.

Detection Method 4 – Detecting malicious ADFS trust modification

An attacker gaining administrative access to ADFS may, instead of extracting the certificate and private key for a standard Golden SAML attack, add a new trusted ADFS. This approach will enable them to sign valid SAML responses and gain persistent access to resources, while evading detection by methods 1-3 . This attack can be detected by monitoring the creation of new ADFS trust, using these events in the ADFS server Windows Security Event Log:

- Event id 307 – "The Federation Service configuration was changed". This event can be correlated to relevant event 510 with the same Instance ID for change details.
- Event id 510 with the same Instance ID – could be more than one event per single 307 event. These events should be reviewed, specifically searching for "Configuration: Type: IssuanceAuthority" where "Property Value" references an unfamiliar Domain.

Detecting SAML attack scenarios in additional architectures

Different SAML architectures involving multiple identity providers and Cloud services may change the way certain SAML attacks are carried out. While these architectures may vary significantly, we decided to focus our attention in this section on two main potential architectures: Cloud identity providers and multiple/hybrid identity provider architectures.

- **Multiple Identity Provider Architectures**
  As different architectures may utilize multiple ADFS instances from different Domains or organizations, it is vital to ensure log collection and analysis is performed on all relevant ADFS. When utilizing a multiple-ADFS or hybrid ADFS-Azure AD architecture, different tokens are created by each identity provider. Attackers targeting this infrastructure can compromise one identity provider, leaving no trace of malicious activity on the other. This means legitimate authentication events may be viewed on the ADFS or Azure AD instance authenticating with the service provider, but not on all upstream instances which may have been attacked, as can be seen in **Figure 3**. To ensure effective detection, each ADFS trust relationship should be analyzed as an IDP-SP pair, following the same logic applied in the previous section (detection methods 1-3).
- **Cloud Only Identity Provider Architectures**
  Federation architecture only utilizing Azure AD as an identity provider mitigates the standard Golden SAML attack vector, as the certificate and private key are stored in Azure instead of in ADFS on-prem. However, an attacker gaining administrative access to Azure AD may add federated ADFS instances under their control, thereby enabling themselves to sign valid SAML responses and gain persistent access to resources. While not a classic Golden SAML attack, this attack can be detected by monitoring the "Set federation settings on domain" activity type and ensuring no unknown ADFS instances are federated to Azure AD. It is also recommended to verify existing federation under Azure AD Connect, to detect previous compromise.

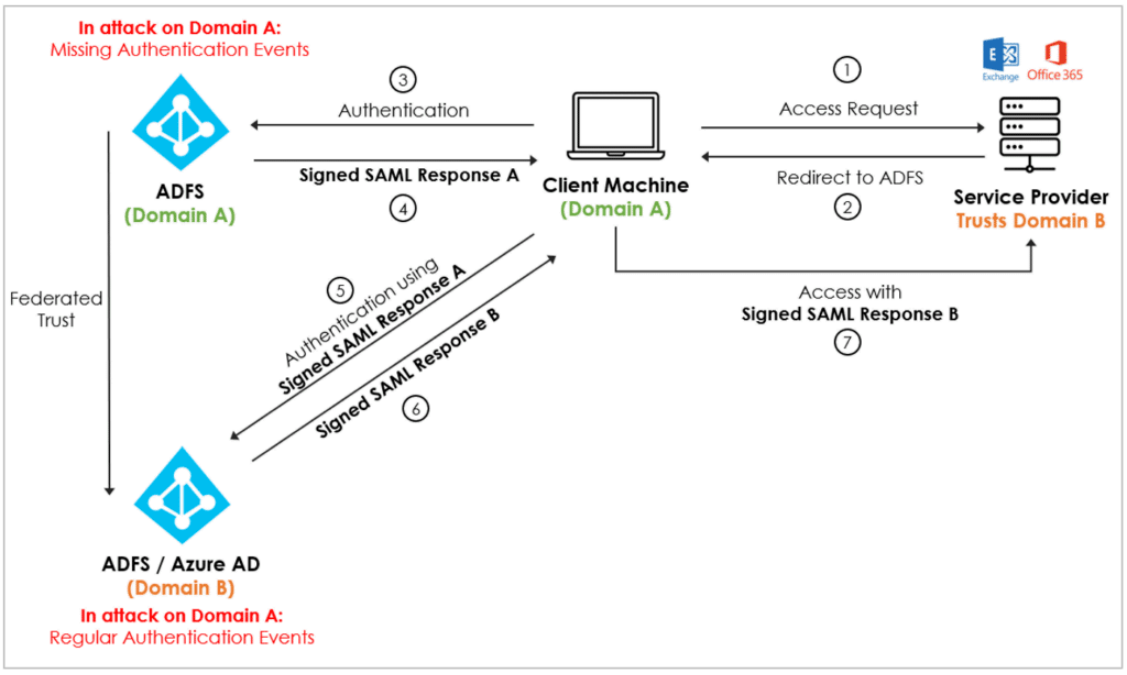

Figure 3: Multi Identity Provider SAML Architecture

[1] CISA Alert (AA20-352A)Advanced Persistent Threat Compromise of Government Agencies, Critical Infrastructure, and Private Sector Organizations.
