July 2021

USING A SECURE TELEPHONY IDENTITY REVISITED (STIR) PERSONAL ASSERTION TOKEN EXTENSION TO IMPROVE AUDIO DIAL-IN EXPERIENCE IN MEETINGS

Ram Mohan R

Rajarshee Dhar

Deepesh Arora

Radha Krishna Saragadam

USING A SECURE TELEPHONY IDENTITY REVISITED (STIR) PERSONAL ASSERTION TOKEN EXTENSION TO IMPROVE AUDIO DIAL-IN EXPERIENCE IN MEETINGS

AUTHORS:
Ram Mohan R
Rajarshee Dhar
Deepesh Arora
Radha Krishna Saragadam

ABSTRACT

Public-Switched Telephone Network (PSTN) dial-in is a common way to join a teleconference (e.g., a teleconference meeting, etc.), which could be an audio-only teleconference or a video teleconference in which the audio portion of the call involves a PSTN path). Many teleconference vendors provide toll-free numbers and/or other PSTN numbers that can be used to dial-in to a teleconference. However, PSTN call-in processes for teleconferences have not evolved much in the last decade and often experience problems. This proposal seeks to address various problems/challenges associated with current PSTN call-in processes for teleconferences by providing novel techniques in which an extension is provided to the Secure Telephony Identity Revisited (STIR) Personal Assertion Token (PASSporT) JavaScript Object Notation (JSON) Web Token (JWT) that includes a new claims section that can be encrypted and secured end-to-end (e2e) such that only an enterprise and/or a teleconference cloud can view the claims section, which can help a meeting cloud to learn an enterprise identity of a participant, can be used to enhance call experience while joining meetings, and/or can help users move from enterprise phones to mobile phones during meetings (e.g., switching audio).

DETAILED DESCRIPTION

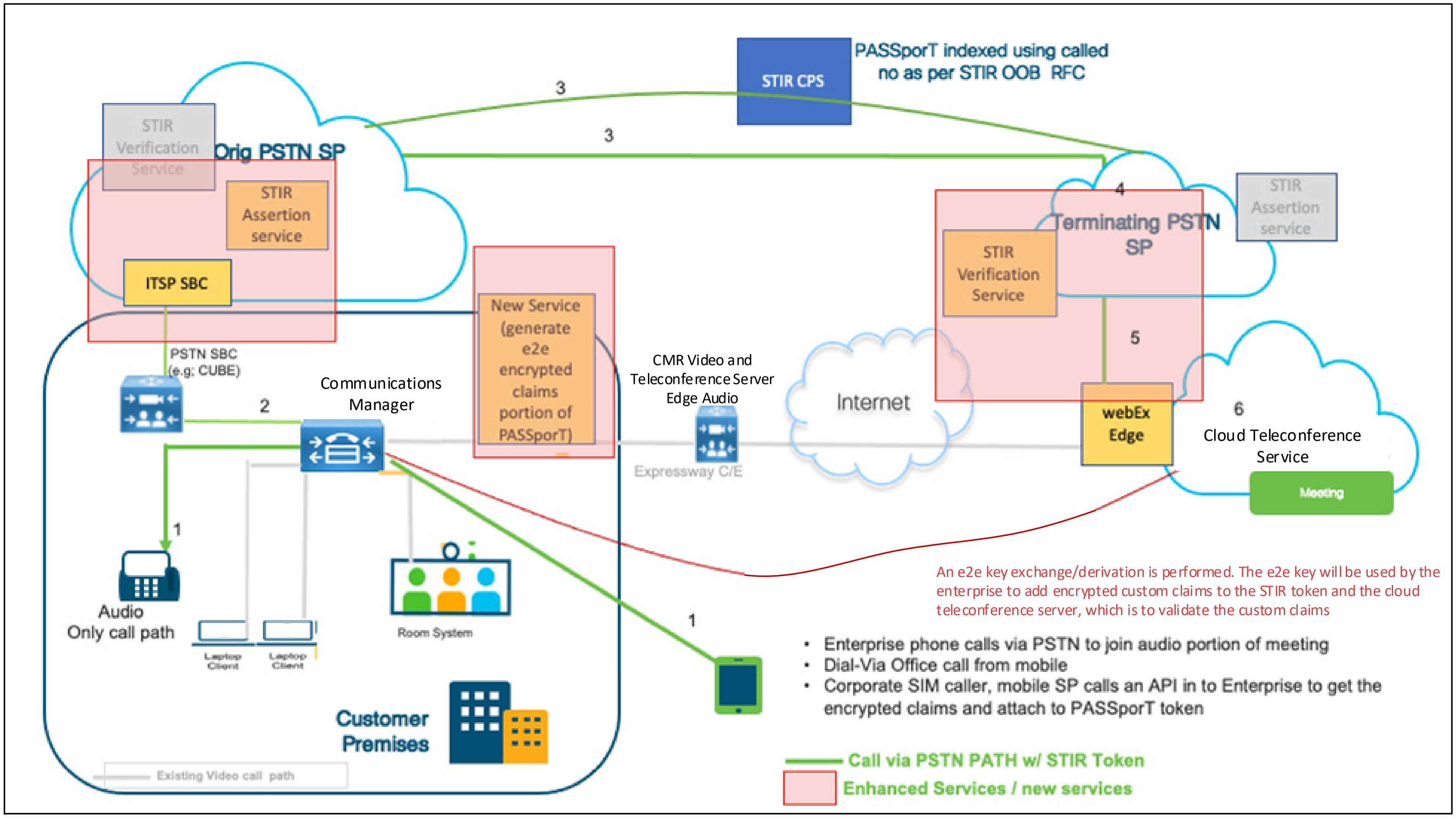
As noted, many audio-only/video teleconferences rely on toll-free or PSTN numbers that can be used to dial-in to a teleconference. For example, up to 40-50% of participants in teleconferences have been found to utilize dial-in audio-only calls, of which, the majority of audio dial-in users can dial-in via a PSTN path either using a corporate/enterprise device or a personal phones, which may or may not have a corporate/enterprise connection.

However, PSTN call-in processes for teleconferences have not evolved much in the last decade and often experience problems. For example, a meeting roster may not reflect participant information accurately, which can cause related analytics to be improper. Further, joining a meeting via PSTN typically involves entering a meeting identity, password, etc. each time, even if a participant is re-joining a same meeting. Still further, if a participant dials in via PSTN for the audio portion of an already joined meeting, the participant still has to enter all the details previously supplied; essentially, the meeting system does not have a way to relate an incoming Session Initiation Protocol (SIP) INVITE from a PSTN path as part of an already established session. Further, it is not currently possible to pass data across a PSTN SIP path securely e2e without various Session Border Controllers (SBCs) looking at the data.

Such problems can lead to sub-optimal experiences when participants want to use different devices to attend teleconferences (e.g., a PSTN path for audio dial-in, an application for sharing information, etc.) and/or for audio-only teleconferences.

This proposal seeks to address various problems/challenges associated with current PSTN call-in processes for teleconferences by providing novel techniques in which an extension is provided to the Secure Telephony Identity Revisited (STIR) Personal Assertion Token (PASSporT) JavaScript Object Notation (JSON) Web Token (JWT) that includes a new claims section that can be encrypted and secured end-to-end (e2e) such that only an enterprise and/or a teleconference cloud can view the claims section. Techniques herein may leverage the STIR/SHAKEN (Signature-based Handling of Asserted information using toKENs) framework (as it is enforced globally) to provide the extension to the STIR PASSporT JWT token, which can help a meeting cloud to learn an enterprise identity of a participant, can be used to enhance call experience while joining meetings, and/or can help users move from enterprise phones to mobile phones during meetings (e.g., switching audio).

Consider an example architecture, as shown below in Figure 1, which illustrates various elements that can be utilized to facilitate techniques of this proposal.


STIR CPS
PASSporT indexed using called no as per STIR OOB RFC
3
3
STIR Verification Service
Orig PSTN SP
STIR Assertion service
ITSP SBC
PSTN SBC (e.g; CUBE)
2
Communications Manager
New Service (generate e2e encrypted claims portion of PASSporT)
CMR Video and Teleconference Server Edge Audio
Expressway C/E
Internet
4
Terminating PSTN SP
STIR Assertion service
STIR Verification Service
5
webEx Edge
6
Cloud Teleconference Service
Meeting
1
Audio Only call path
Laptop Client
Laptop Client
Room System
1
Customer Premises
Existing Video call path
Call via PSTN PATH w/ STIR Token
Enhanced Services / new services

An e2e key exchange/derivation is performed. The e2e key will be used by the enterprise to add encrypted custom claims to the STIR token and the cloud teleconference server, which is to validate the custom claims

- Enterprise phone calls via PSTN to join audio portion of meeting
- Dial-Via Office call from mobile
- Corporate SIM caller, mobile SP calls an API in to Enterprise to get the encrypted claims and attach to PASSporT token

Figure 1: Example Architecture

Consider, in the context of Figure 1, various example use-cases in which techniques herein may be utilized to enhance/improve user experience (UX) for audio dial-in calls for teleconferences:

1. A call originating from corporate/enterprise Voice-over-Internet Protocol (VOIP) phone (e.g., registered to enterprise call control) in which a participant dials-in to a teleconference bridge and the call is routed via a PSTN SIP edge of the enterprise as shown in Figure 1, in which:
    a. The participant tries to join a meeting from a primary desk phone; and/or
    b. The participants logs-in into any desk phone available using Extension mobility (e.g., the participant is still inside the enterprise network);
2. A call originates from a cell phone in which the call is routed via an enterprise Dial-Via Office (DVO), as shown in Figure 1;
3. A call originates from a cell phone to a teleconference bridge but the caller's Subscriber Identity Module (SIM) is an enterprise-issued connection/number;
4. A call originates from a traditional PSTN mobile phone (e.g. without an enterprise SIM card /connection) with no VOIP connection. Imagine, for example, that the user is outside the enterprise network and doesn't have an enterprise SIM card but wants to join a meeting; and/or
5. In addition to these, the techniques of this proposal can also be used to improve the UX of a call switching from VOIP to PSTN. For example:
    5.1 If the VOIP (over the top path from enterprise to cloud) media path quality is bad, there are techniques that allow the call to switch to PSTN path or a teleconference participant can use PSTN path for the audio portion of the teleconference. When this occurs, instead of the participant being required to provide a meeting identity (ID) again, the techniques of this proposal can be used to send a session identifier that can help the cloud teleconference service recognize that the participant is already participating in the setup call.

In some instances for use-cases, 1-3 and 5, the enterprise phones can maintain a calendar service (e.g., once the user logs-in), based on in which the calendar service can fetch the teleconference details (e.g., teleconference/meeting (ID), access code, etc.) and later use the meeting details with the STIR/SHAKEN framework in order to join dial-in teleconferences automatically.

Whereas, for use-case 4, the enterprise phones can reach out to the user's personal PSTN phone (VOIP or Not VOIP) in order to remind the user about an ongoing teleconference (e.g., by using a calendar service, as explained above). In one instance, the user can be presented with an Interactive Voice Response (IVR) service that indicates, as an example, "You have a Meeting XXXX, Press 1 to join here." In this example, one the user confirms (e.g., by pressing 1), the enterprise phone first joins the dial-in bridge with the meeting details (as explained above) and then transfers the PSTN phone the teleconference bridge.

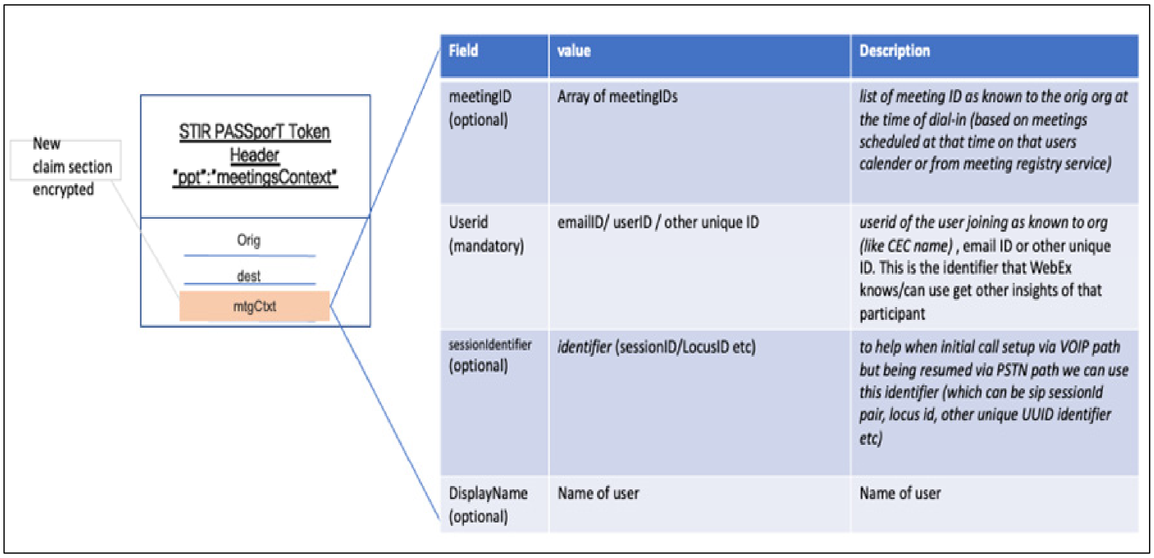
Here, before transferring, the phone can notify the teleconference dial-in bridge regarding a call transfer that is going to happen (e.g., using additional custom fields in a STIR token, beyond those discussed), so that the teleconference dial-in bridge can authenticate the PSTN user and patch the user in place of the enterprise phone.

Although various examples provided herein are discussed with reference to an audio teleconference, it is to be understood that the techniques of this proposal can be equally applied to video teleconferences in which the audio portion of the teleconference involves PSTN dial-in at some point during a teleconference workflow.

Based on the architecture as illustrated in Figure 1, consider various details of the techniques of this proposal. STIR is a framework that enables the cryptographic assertion and verification of the identity of a caller. While the STIR framework allows a caller/authentication server to cryptographically assert a caller identity via the SIP Identity header field in a SIP INVITE, existing deployment realities would make it difficult if not impossible for the Identity header field value to be preserved end-to-end (e2e) from caller to callee. This is in large part due to the significant deployment trail of legacy equipment in carrier networks.

Taking these deployment realities into consideration, STIR suggests an out-of-band mechanism, as can prescribed at: https://datatracker.ietf.org/doc/draft-ietf-stir-oob/ in which an entity that cryptographically asserts caller identity (authentication service) places a full PASSporT [as prescribed by Internet Engineering Task Force (IETF) Request For Comments (RFC) 8255] in a rendezvous service referred to as a Call Placement Service (CPS). To verify the identity of the caller, the verification service running at the callee contacts the same CPS to obtain, decrypt, and verify the PASSporT identity assertion.

Consider various operations or steps that can be performed to implement the techniques of this proposal. As a first step (1), an extension to the STIR PASSporT token is proposed that can be provided by inserting new "claims" in the token that are e2e encrypted. This new claim will carry additional information, as shown below in Figure 2.



Field	value	Description
meetingID (optional)	Array of meetingIDs	*list of meeting ID as known to the orig org at the time of dial-in (based on meetings scheduled at that time on that users calender or from meeting registry service)*
Userid (mandatory)	emailID/ userID / other unique ID	*userid of the user joining as known to org (like CEC name)* , email ID or other unique ID. This is the identifier that WebEx knows/can use get other insights of that participant
sessionIdentifier (optional)	*identifier* (sessionID/LocusID etc)	*to help when initial call setup via VOIP path but being resumed via PSTN path we can use this identifier (which can be sip sessionId pair, locus id, other unique UUID identifier etc)*
DisplayName (optional)	Name of user	Name of user

Figure 2: New Claim Section Added to a STIR PASSporT Token

For a second step (2), RFC 8225, Section 8 describes that a PASSporT JWT can be extended using any protocol. Thus, the JWT header of a PASSporT token can be enhanced to include a "ppt" type "meetingsContext" as per the RFC 8225, Section 8.01 recommendation, as follows:

```
{
"alg":"ES256",
"ppt":"meetingsContext",
"typ":"passport",
"x5u":"https://tel.example.org/passport.cer"
}
```

For a third step (3), a new meeting context extension claim called "mtgCtxt" is proposed that can be added on top of the base JWT claims ("Orig," "dest," and "mky") as defined in RFC 8225. The various fields are described in Figure 2, above. Below is an example instance of how the extension may look to a user who has one meeting scheduled at the time at which the user is trying to join a teleconference:

```
"mtgCtxt": {
"meetingID": [123456789]
"userid": "User1@enterprise.com"
"DisplayName": User1
"accessCode" : 123456
}
```

If the INVITE sent out is for an already setup call (e.g., a participant uses a teleconferencing application to join but audio is provided via PSTN or the participant initially calls-in using a computer but switches audio mid-way), then in such cases additional context information can be carried that will carry a session identifier of an already setup session. Further, in cases where a transfer is being performed, the original STIR can be referenced so that the cloud teleconferencing server can determine that the incoming INVITE from PSTN path matches the previously created session, as follows:

```
{
TransferToPSTN: True
Transferred details : {
TranferredBy: 55515251617
TransferredByUser: User2@enterprise.com
TransferredByName: User2
JoinAutomatically: True // optional for cases where user is already authenticated can be used depending on org security policy
Authenticated:True
}
}
```

For a fourth step (4), the new claim can be a JSON portion that is e2e encrypted using a key that is shared between the enterprise and the cloud teleconference service. Various techniques may be utilized to facilitate the e2e key derivation, as follows:

- The enterprise and the cloud teleconference service can exchange the verification keys out-of-band (e.g., via a Hypertext Transfer Protocol (HTTP) channel or other similar means). Given that the enterprise and cloud teleconference service already will have some association (e.g., a site admin account, etc.), the exchange can be performed using a public/private key pair to exchange the e2e key on a secure connection between the enterprise and the cloud teleconference service;
- An enterprise administrator, via a continuous integration (CI) token, can create and push the e2e keys to the cloud by calling a secure Application Programming Interface (API) that the cloud teleconference service can expose. For example:

  ```
  POST https:/teleconf.service.com/api/v1/e2eKeys
  {OrgID, e2eKey, e2eKeyIdentifier}
  ```
- The e2e key can be derived by exchanging a Key Derivation Function (KDF) 'salt' using a secret that is pre-configured on-site for the enterprise administrator account and is known to both the enterprise and the cloud teleconference service. It should be noted that there can more than one key exchanged along with a "kid" between the enterprise and the cloud teleconference service in this example.

For a fifth step (5), the "mtgCtxt" claims in the STIR PASSporT will be encrypted using one of the keys, derived as described above, whose "kid" will be conveyed in the

JSON Web Encryption (JWE) header of the JWT. The encrypted JSON claims would have a JWE header with a kid that will help the destination side fetch right keying material to decrypt the encrypted portion of the JWT. The JWE header will be base64 encoded and the "mtgCtxt" overall will be generated using procedures as prescribed by RFC 7516.

An example JWE can be formatted, as follows:

```
{
"alg": "dir"
"cty": "JWT",
"enc": "A128CBC-HS256",
"kid":"a21f862a-fd5d-47cb-8e06-
        978ccace8486:890a314f3dba07119e67acecc4eccd6949de5743724012bef5
        0817edca4ad97c"
}
```

For a sixth step (6), the encrypted claim will be nested inside the PASSporT token as a new claim and sent in INVITE via an Internet Telephony Service Path (ITSP) path/PSTN path. The existing STIR procedures of RFC 8225 will be applied to sign the entire STIR token, which will also include the "mtgCtxt" claim section.

For a seventh step (7), consider that typically, for a PSTN call, the ITSP provider (e.g., service provider) will generate a STIR token. The service provider will likely have infrastructure or partners to generate the token. In some instances, a PSTN service within call manager cloud can for enterprise by partnering with STIR/SHAKEN vendors.

Various potential flows are possible for the seventh step. Figure 3, below, illustrates an example flow in which an ITSP provider adds the STIR token on the outbound INVITEs obtained from an enterprise. For example, an enterprise phone can make the call that gets routed via the PSTN ITSP. In this case, one implementation may just allow the ITSP perform the STIR token creation. For the flow, the enterprise will pass the "mtgCtxt" in a custom SIP header of an outgoing INVITE to its partner ITSP provider. The ITSP partner will create STIR token and while doing so, pick up the "mtgCtxt" from the SIP header passed by the enterprise and add the "mtgCtxt" while creating the PASSporT. Given that "mtgCtxt" is encrypted, the ITSP will not be able to tamper with the "mtgCtxt" blob.

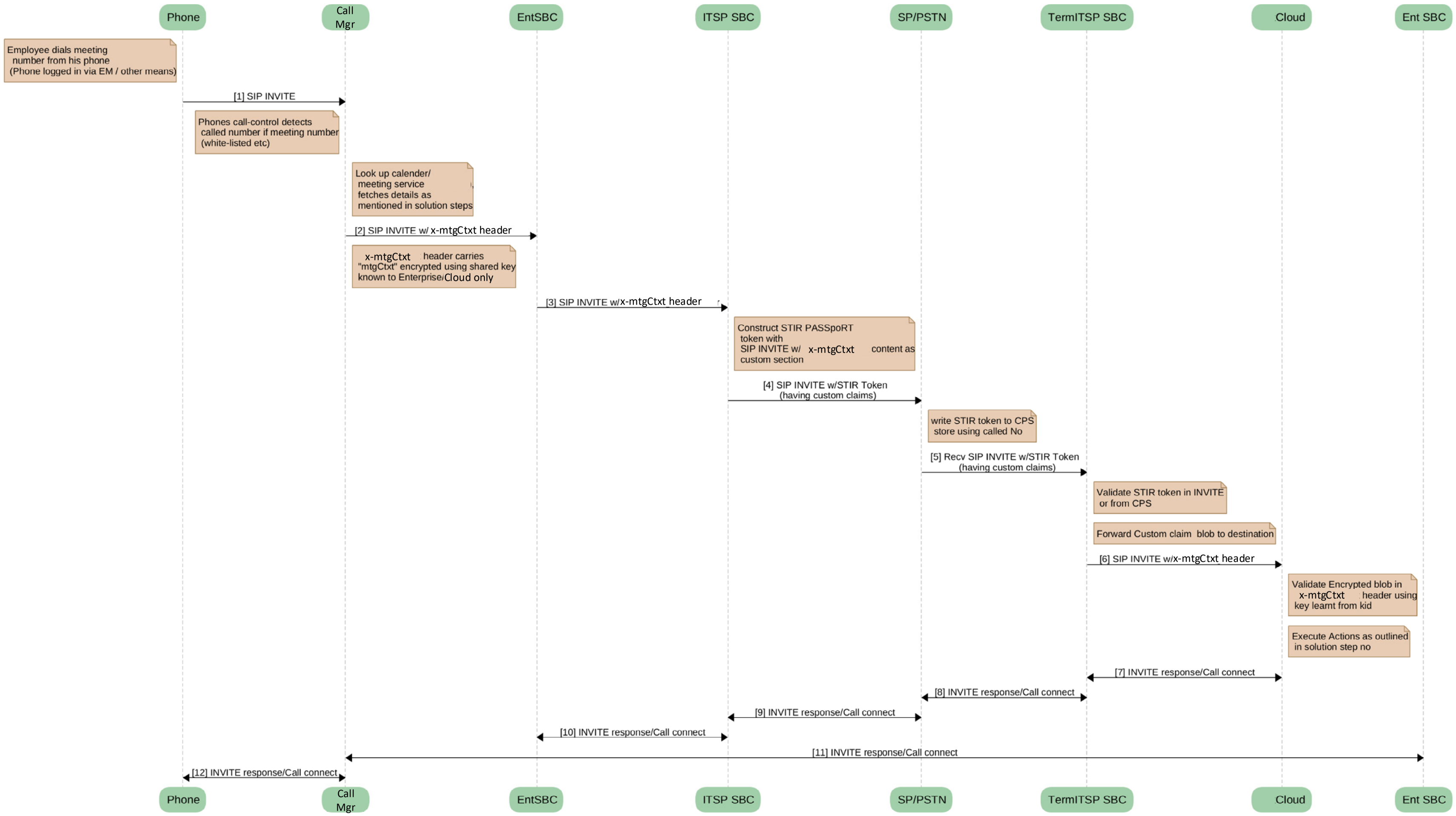

Phone
Call Mgr
EntSBC
ITSP SBC
SP/PSTN
TermITSP SBC
Cloud
Ent SBC
Employee dials meeting number from his phone (Phone logged in via EM / other means)
[1] SIP INVITE
Phones call-control detects called number if meeting number (white-listed etc)
Look up calender/ meeting service fetches details as mentioned in solution steps
[2] SIP INVITE w/ x-mtgCtxt header
x-mtgCtxt header carries "mtgCtxt" encrypted using shared key known to Enterprise/Cloud only
[3] SIP INVITE w/x-mtgCtxt header
Construct STIR PASSpoRT token with SIP INVITE w/ x-mtgCtxt content as custom section
[4] SIP INVITE w/STIR Token (having custom claims)
write STIR token to CPS store using called No
[5] Recv SIP INVITE w/STIR Token (having custom claims)
Validate STIR token in INVITE or from CPS
Forward Custom claim blob to destination
[6] SIP INVITE w/x-mtgCtxt header
Validate Encrypted blob in x-mtgCtxt header using key learnt from kid
Execute Actions as outlined in solution step no
[7] INVITE response/Call connect
[8] INVITE response/Call connect
[9] INVITE response/Call connect
[10] INVITE response/Call connect
[11] INVITE response/Call connect
[12] INVITE response/Call connect
Phone
Call Mgr
EntSBC
ITSP SBC
SP/PSTN
TermITSP SBC
Cloud
Ent SBC

Figure 3: Example Call Flow

Various key steps illustrated in the call flow include that enterprise will generate encrypted claims using the procedures of RFC 7516. The claims can various content as discussed above for the first step and will be added to the x-mtgCtxt SIP header. An example header may be formatted as, x-mtgCtxt: token <value>, where <value> is an RFC 7516 encoded token (whose header is base64 encoded and can provide the kid to use). Next, the ITSP attaches this as a new claim to STIR PASSporT and signs the entire passport following procedures of RFC 8225.

Optionally, the STIR token can be placed in CPS and fetched using the called number as per existing procedures. In different implementations, a CPS service can be implemented by service providers, third-party providers, a cloud telephony provider that also provides a PSTN break out, or even facilitated through partnerships among different providers. Thus, three CPS models may be possible. A first model may include service providers owning a CPS service and an originating service provider (e.g. that is connected to the enterprise for the flow of Figure 3) will sign/generate the STIR and a terminating SP (that connects to cloud teleconference service edge) will validate the STIR token with using CPS (whenever e2e SIP is not involved). A second model may include a third-party owning a CPS service and partnering with an enterprise. A third model may include a cloud provider owning a CPS service (e.g., a PSTN service or PSTN edge that can own/operate a signing authority over many numbers owned by the provider).

Continuing with the present flow, at the terminating side, the termITSP validates the STIR token, picks up the "mtgCtxt" portion, and sends this portion in the INVITE x-mtgCtxt SIP header to the cloud teleconference service. Thereafter, the cloud teleconference service performs a base64 decode, obtains the kid from the header, fetches the appropriate keying material of the corresponding enterprise organization ID (OrgID) and decrypts the "mtgCtxt" portion.

In cases where the call is originated by a mobile device (e.g., use-cases 2 and 3 above), the ITSP (if it happens to be a partner of the cloud teleconference service) can detect that the called number is a teleconference toll-free number/meeting number and can call an API into a service exposed by enterprise to determine the encrypted blob.

For an eighth step (8, continuing from the seventh step, noted above), consider that on the termination side (e.g., cloud teleconference service edge where the server peers with the terminating ITSP), the ITSP validates the STIR token and will extract the claims portion (e.g., JWE encrypted claims), inserts a new X-header (e.g., x-mtgCtxt), and can send the same in the forwarded INVITE from ITSP to the cloud teleconference service edge can simply forward the PASSporT to the edge and have the edge itself validate the STIR token (e.g., a delegated validation).

Thus, techniques proposed herein can be provided in-line as extensions to the standard RFC for the STIR PASSporT. The STIR token will be carried e2e, from the perspective of an originating service provider to a terminating service provider and the existing relationship between the service providers and an enterprise/cloud teleconference service will ensure that data is forwarded e2e. Middle boxes can validate any known extensions (e.g., any standardized extensions) and pass across the token and any other claims that are not understood. The new claims proposed herein are e2e encrypted for the purpose of protecting the information from being viewed/tampered with by any PSTNs/service providers.

Accordingly, various teleconference/meeting experience improvements once the edge validates the encrypted claim (after decrypting with appropriate keying material) can be provided. For example, currently when a user joins a teleconference via a PSTN path, the user's enterprise identity cannot be asserted. However, the "mtgCtxt" claim of this proposal can carry the enterprise identity of a user, which can help to solve this problem. Further, currently when a user joins a teleconference via a meeting application or other device but uses a phone to dial-in to the audio portion of teleconference, there will be two users shown in roster/participant list. This proposal provides for solving this problem in order to facilitate relating the two users by means of correlating a "mtgCtxt" session identifier and an enterprise identity of the user.

Additionally, in some instances, the service that plays teleconference/meeting prompts (e.g., a GIVR service) can use the encrypted claims as a first level of authentication (as the encrypted claims are signed using a key known only to the enterprise and the cloud teleconference service) and can directly progress to the step at which it can validate a second factor of a given participant (e.g., personal identification number (PIN), password, voice fingerprint, etc.).

Further, the display name in the roster of a teleconference/meeting, once an attendee has joined, can reflect the actual name of the attendee as fetched from the "mtgCtxt" claim. Additionally, since the cloud teleconference service is able to verify the identity of a participant that has joined via a PSTN path, the cloud server can use this identity to fetch further information for the participant, such as a people insight profile, social graph of the participant, etc.

In the case of a call being switched from VOIP to PSTN (e.g., if a participant initially joined via a VOIP path), then the mtgCtxt claim can also carry a call reference and/or call identifier of the earlier session in order to help stitch/relate to the previous session and, hence, send the media back to the same media bridge, optimize resource allocation, etc.

Still further, the "mtgCtxt" can also carry keying material for cases in which e2e encryption for audio is desired. Since "mtgCtxt" is carried securely over PSTN path, this can be used to carry any keying material and then can be used to exchange/derive actual Secure Real-Time Transport Protocol (SRTP) keys in the media path.

As noted, CPS is primarily a service intended to carry STIR tokens out-of-band. Thus, it is possible to de-couple the enhanced token/context information of this proposal and not carry it in PASSporT. In one instance, to decouple the enhanced token from a STIR/PASSporT token, a third-party/cloud CPS service provider can offer a new service in which an additional entry can be provided in a CPS table row to store this context (and potentially offer this to customers, for example, via PSTN offering). Thereafter, the context can be used by an enterprise in order to push the additional meeting context whose identifier could be a calling/called number or any other existing unique identifier (e.g., meeting ID, etc.) in order to correlate an incoming SIP INVITE from a PSTN service with the context in order to fetch the context from the table. In such a case, the enterprise can avail this service to store the additional context information (here "mtgCtxt") and have the cloud teleconference service edge fetch the context when the incoming SIP call hits the edge from PSTN.

In summary, provided herein are techniques through which additional context information via a PSTN SIP path from an originating network (e.g., an enterprise/cloud teleconference service) to a terminating network (e.g., an enterprise/cloud teleconference service, depending on the direction of the call. The additional context information (e.g., "mtgCtxt") can include an organization identity of a user, meeting/teleconference details (where possible), and e2e keying material that can be used to derive e2e keys (e.g., for cases in which Secure Media Frames (SFRAME) may be used). Thus, various problems/challenges of current implementations can be solved by the novel techniques of this proposal in which an extension is provided to the STIR PASSporT JWT that includes a new claims section that can be encrypted and secured e2e such that only an enterprise and/or a teleconference cloud can view the claims section.