Charging of SAML-based Federated VoIP Services

David J. Lutz, Dominik Lamp, Patrick Mandic
Universitaet Stuttgart
Rechenzentrum Universitaet Stuttgart
Allmandring 30a, Stuttgart, Germany
Email: {lutz, lamp, mandic}@rus.uni-stuttgart.de

Fabio Hecht, Burkhard Stiller
University of Zurich
Department of Informatics (IFI)
Binzmühlestr. 14, 8050 Zürich, Switzerland
Email: {hecht, stiller}@ifi.uzh.ch

Abstract

Whilst SAML-based federations are most often used by academic and semi-commercial institutions that focus only on attribute-based authentication, we foresee a growing interest for service providers providing charged services. Since more and more academic and semi-commercial federation participants offer Voice-over-IP (VoIP) services, this type of service provides an entry point into identity federation based payment. Therefore, this paper describes an approach on how to harmonize the SAML-based federation technology with the needs of a payment infrastructure for enabling charging of VoIP services within a federation. However, since the different aspects of our approach (SAML Payment, SIP Discovery and Tariff Function) are not bound to VoIP applications, each one of them could be used separately or combined for several service types.

1 Introduction

SAML [1] based federations were often deployed to avoid users having to have a separate account at each Service Provider (SP). Bundling all identity information at one Identity Provider (IdP) and transmitting the required information to requesting SPs allows for a comfortable identity and access management within the federation. However, until now, federations have focused mainly on authentication and authorization. Now, however, as more and more SPs intend to join such federations, another issue must be managed: payment. Since most federations rely on SAML, this paper provides a solution for integrating SAML-based federation technology (defining SAML assertions for payments) with a payment infrastructure to enable charging for VoIP services within such a federation.

One of the first charged services that SPs can benefit from are Voice-over-IP services, as already deployed by universities or semi-commercial institutions. Although these services are often free-of-charge when using the IP network, charging comes into play when the callee’s number is located outside the IP network. When considering an international federation, such as [2], VoIP SPs may choose another federated SP to route the call to the commercial telephone network, e.g., because the SP is located within the same country as the called number. In this case, the foreign VoIP SP (FSP) has to charge the user’s home VoIP SP (HSP) for the outgoing call and the HSP has to forward all occurring charges to the user. For finding the best fitting VoIP SP, a newly developed discovery service may be chosen and the charging can be done using an innovative charging mechanism. However, payment for the call is handled using the above mentioned SAML payment approach.

A typical scenario, as foreseen for the approach described may be viewed as follows: A student (S) has an account at his university that offers VoIP-calls that are charged when being routed through the commercial phone network (PSTN). When S visits a conference abroad, he wishes to make a call from his hotel to a local number. However, when he tries to set up the connection, his home VoIP SP (HSP) checks whether there are other VoIP SPs within the federation offering a cheaper connection. HSP uses the proposed discovery mechanism and finds a foreign VoIP SP (FSP) close to S’s location. Thus, HSP initializes the call from S via FSP to the PSTN network. For the cost claims between HSP and S, HSP uses a special tariff function to compute the costs online and charges S for the costs, who then pays using SAML Payment Assertions.

We see a separated audience for this approach: federated VoIP SPs could extract the whole idea described here, whilst other SPs as well as other federation members could gain benefits by deploying only some aspects mentioned, such as payment or discovery or tariff functionalities. However, it needs to be mentioned that the payment approach does not compete with the already deployed payment mechanisms of commercial service providers; its goal is to provide reliable payment mechanisms within the federation software to SPs who have not, as yet, implemented a payment mechanism.

In order to describe this new approach for SAML-based
charging of federated VoIP-Services, the architecture is first described, covering the proposed VoIP networks and SAML federations. Next, an idea on how the best VoIP service can be discovered within the federation is introduced. Then, an innovative mechanism to adequately charge for VoIP traffic is explained, followed by further sections that consider security and implementation issues. The paper concludes with a summary of the results achieved.

2 Related Work

The analysis of related work has been split into the area of Identity Federations and VoIP Charging, which are both explained in the following subsections.

2.1 Identity Federation

The Security Assertion Markup Language (SAML) [1] is currently being used by many identity federations as a reliable language for transmitting authentication and authorization data as well as the required federation participant’s attributes. Shibboleth [3] has been deployed in an academic context, while, with Liberty Alliance [4], a federation infrastructure approach has been designed for a business audience. Within such a federation, several service providers agree on having user authentication not on their systems, but at the user’s home institution, e.g., the user’s university or his/her telecom operator. This home institution, the so-called Identity Provider, furnishes, upon request, authentication and attribute data about a specific user to federated service providers. Therefore, these service providers are able to check, based on the received user’s data, whether the user is allowed to have access to a specific resource or to consume a specific service.

2.2 VoIP Charging

One of the main enablers of platforms such as IMS is the ability to charge the user for the services provided, as was the case for the old PSTN and the 2/3G architecture. To this end, these old mechanisms have been mimicked in the current IP-based architecture. As in non-IP networks, both online (pre-paid) and offline charging exist in most IP architectures [5],[6]. Most notoriously, the IMS uses the IETF Diameter protocol for charging events. In the offline mode, there are two types of charging methods, the event-based charging and the session-based charging whereas in the online mode three methods can be distinguished: the immediate event charging (IEC), the event charging with unit reservation (ECUR) and the session charging with unit reservation (SCUR). In IEC, before the service takes place, a number of credit units are subtracted from the user’s account and, if not enough resource units are available, the service is denied. In ECUR, some credit units are reserved in the user’s account and after the service has been executed, the resources used are committed. In the last case, the SCUR, an initial unit reservation is performed and updates are sent subsequently during the session reporting the units used and requesting additional units. In order to obtain the rate of a certain service applied to a certain user, the so-called “Rating Function” is used.

3 Architecture Analysis and Design

Within the scope of this work, two different architectures need to be analyzed: the VoIP architecture for the network and the architecture of the enhanced SAML federation.

3.1 VoIP Network

In traditional (2G) networks for user-to-user communication, roaming mechanisms are established at the network level. Thus, after discovering a 2G network that allows the roaming user to place calls, both in and outbound calls can be routed through the visited network’s infrastructure.

In contrast, transport and voice-service networks, as usually deployed by, e.g., academic institutions, are distinct infrastructures. Network access is granted on an individual or federation-based agreement. Provided that the effective network access policies allow a user to connect to his/her home network infrastructure, the network provider is only required to act as a bit pipe.

To remain involved in service delivery, including audio and video calls, the (mainly telco) operators developed the Internet Multimedia Subsystem (IMS) Service Platform [7] for service provisioning. Different Call Service Control Functions (CSCF) are deployed to deal with user authentication and session handling. Although our approach is contrary to IMS and the network provider is reduced to a bit pipe, the canonical architectures are quite similar. Thus, where appropriate, the well-established IMS terminology is used in this paper.

When roaming, a user may continue to use his/her home VoIP infrastructure and, therefore, to use his/her geographical home number. This, however, has two disadvantages: Firstly, calls to the guest institution are routed over the PSTN and are potentially charged at higher rates compared to a local breakout. Secondly, additional latency is introduced into calls to the guest institution and the surrounding area. According to [8], the mouth-to-ear delay should not exceed 150ms.

For the above reasons, we propose to use a local breakout while staying independent of the actual access network. In Figure 1, the basic architecture is depicted: The user has network connectivity to his/her home institution and registers with his/her home institution’s Proxy-CSCF (P-CSCF).
The home P-CSCF detects when it is cheaper to route a call through the local IP-PSTN gateway. If not, the user’s client is referred to the Session Call Service Control Function (S-CSCF) server at the remote institution. The client then places the call via the remote institution’s IP-PSTN gateway to the targeted phone within the PSTN.

Charging is carried out on the user’s Payment Provider account. Thus, the participating S-CSCF and P-CSCF servers perform online charging, i.e., they propagate charges as soon as they occur. This process is similar to standard billing and charging processes in the IMS, but instead of accounting messages and money reservation, SAML tokens representing virtual money are exchanged. This novel token-based approach is presented in detail in sections 3.2.6 and 5.

3.2 SAML Federation

It is assumed that the federation described here follows the ideas of the major SAML-based federations, such as Shibboleth [3] and Liberty Alliance [4]. This means that, when using SAML, the three typical components: User, IdP and SP are deployed. However, conventional SAML-based federations do not focus on payment solutions. Thus, a new component has to be introduced to handle payment within the federation: the Payment Provider (PP).

3.2.1 User

The User is assumed to be either a human or an automated program that needs to access a protected resource at an SP. Regarding the philosophy of SAML-based federations, the following two assumptions are made about the User’s behaviour within the payment-enabled federation:

- Account at IdP: the User has an account at one of the federation’s IdPs, i.e., this IdP hosts all the User’s identity data that may become relevant to the federation, such as authentication and attribute information.

- Account at PP: for handling the payment within the federation, the User has chosen one of the federation’s PPs to host his/her payment account. This means, that the User has stored a sum of money at the PP and the PP is allowed to generate payment assertions and to charge his/her account for the assertion-based expended money.

3.2.2 Identity Provider

The Identity Provider (IdP) is the component that is responsible for a reliable handling of the User’s identity data within the federation. It authenticates the User during the initial login, sends authentication assertions to him/her for further authentication steps at the federation SPs and generates SAML Attribute Assertions whenever an SP requests such an assertion and the User’s privacy policy, the so-called Attribute Release Policy, allows this transmission.

3.2.3 Service Provider

The Service Provider (SP) offers service usage and resource consumption to the users within the federation. Such a service could be a simple web-page access or a complex composed service. However, not only the kind of service, but also the kind of access control and access policy can vary. The SP could allow access to (a) everyone, (b) authenticated users, (c) attribute-based authorized users, or (d) paying users. Case (d), being new to federations, requires interactions with the newly developed Payment Provider.

3.2.4 Payment Provider

The Payment Provider (PP) is the new component that does not exist within current federations. It is responsible for handling the payment within the federation in a similar manner as the IdP does for handling the User’s identity information. It has to handle the three payment-related tasks Assertion Generation (upon a User’s request, the PP generates a SAML Payment Assertion for him/her), Assertion Validation (since the payee of an assertion-based payment needs surety about the assertion’s validity, the PP has to check whether the assertion is valid), and Assertion Reimbursement (the PP has to reimburse the assertion into real money when requested). These tasks are reiterated in section 3.2.6 when describing the three main payment processes.

3.2.5 Payment Assertion

The SAML Payment Assertion is a newly developed SAML assertion to transmit payment information securely. To pro-
vide all the necessary information, the SAML Payment Assertion has to carry the following eight pieces of information:

- **Amount**: all participants must know the value conveyed by the assertion.
- **Currency**: all participants must know the currency conveyed by the assertion.
- **PP ID**: the SP must know where to reimburse the payment assertion received.
- **PP Signature**: surety about the PP is needed during reimbursement.
- **Payer ID**: the payer must be identified for account charging or should he/she act maliciously.
- **Payer Signature**: non-repudiation when the payer’s account is debited or should he/she act maliciously.
- **Payment Assertion ID**: needed by the Payment Provider to detect misuse.
- **Payment Assertion Lifetime**: needed to control the assertion’s validity.

Using SAML for transmitting this information brings two advantages: Firstly, SAML offers a high level of security and, secondly, by using assertions the payment process can easily be integrated into the federation’s architecture and its protocols.

### 3.2.6 Payment Processes

Although the payment processes are not part of the architecture but rather of the use-cases, they are described here in order to understand the working methods of the PP. There are three processes that are relevant to the SAML payment approach: assertion generation, assertion payment and assertion reimbursement.

- **Assertion Generation**: The Assertion Generation is initiated by the User who sends an assertion request to his/her PP. Afterwards, the PP performs an authorization step, since the PP must be sure that: (a) the User is who he/she claims to be, (b) the User is authorized to request the assertion, and (c) the User has enough money at his/her PP account. During the assertion generation, the PP stores the assertion’s validity information (such as lifetime and assertion ID) locally for further validation requests.

- **Assertion Payment**: The Assertion Payment starts with an SP-initiated payment information transmission after the User has tried to access a service offered. The User sends back the payment assertion that contains the required amount to the SP (if the User has not received such an assertion from his/her PP, he/she has to request it now). The SP validates the assertion, which means that it first checks the assertion’s amount and currency and, secondly, requests validation at the issuing PP. This validation is done by checking the locally stored assertion’s validity information at the PP. After having received a successful validation message from the PP, the SP informs the User about the successful payment and grants him/her access to the requested service.

- **Assertion Reimbursement**: The third process is the reimbursement of the assertion into real money. For payment within the federation, assertions are used, but when the money leaves the federation, the assertion has to be reimbursed into real money. To do so, the payee of the assertion-based payment sends the received assertion with a request for reimbursement to the issuing PP. The PP checks the assertion by examining the locally stored assertion’s validity information and, if the assertion is valid, the PP credits the requestors account with the assertion’s monetary value.

### 4 Service Discovery

To obtain knowledge about the best fitting VoIP SP within the federation, a service discovery mechanism may be used. The discovery of appropriate federated SPs can be quite dynamic and means need to be developed in order for SPs to publish their information as well as for users to retrieve it.

The retrieval of this information can be performed by a multitude of protocols, each with its own advantages and drawbacks. Regarding identity federations, a suitable approach could have been the use of the Liberty ID-WSF Discovery Service [9]. However, here we tackle a much broader approach in terms of discovery of resources, identities as well as end-user services and payment services. Mechanisms are required for updating the state of services in order to be able to continuously select the most suitable service and to seamlessly swap services during execution or when they may suffer temporal unavailability, e.g., when being delivered from mobile devices. In our case, the chosen protocol is SIP. This reduces complexity as the same protocol is used for information discovery and for establishing VoIP communications, thereby, providing an integrated and generic solution. In addition, this solution can convey simultaneously different kinds of data such as txt, xml, owl just by specifying the content type in the packet. That way, the same service discovery infrastructure can be reused for different purposes.

This proposal relies on the work of the service discovery
mechanism explained in more detail in [10], which is based on two SIP extensions, namely "SIP Specific Event Notification" (RFC 3265) and "SIP extension for Event State Publication" (RFC 3909).

4.1 Modules

From a conceptual point of view, a service discovery architecture is usually comprised of three main actors: a Service Provider, which offers the service, a Service Requestor (the inquiring entity) and a Repository or Directory Agent, which facilitates a requestor to select the SP that best matches its requirements. The definition of the main entities involved and its mapping with the usual service discovery actors is as follows:

- Service Publication Agent (SPA): This is a SIP User Agent supporting SIP PUBLISH which will be in charge of making the service capabilities public.
- Service Discovery Subscriber (SDS): This is the consumer of the information (i.e. the service requestor).
- Service Discovery Agent (SDA): This is the Directory Agent in the service discovery model. This entity is primarily responsible for storing, aggregating and maintaining service information from different SPAs and the corresponding subscriptions from SDSs. Additionally, due to the SIP mechanisms it is based on, it is also capable of informing the subscribers if any change in the services capabilities they are interested in occurs.

4.2 Functional Behaviour

The mechanisms required to carry out the discovery of the desired service providers rely on two basic functionalities:

- Service Publication: This is the way in which a service provider registers and publishes its services in the Directory Agent to be discovered by Service Requestors.
- Services Search: This mechanism defines how a requestor can make a request to the repository, indicating the services features it is interested in, and how the reply is delivered.

These mechanisms are described in Figure 2. The publication mechanism relies on SIP PUBLISH, while a specific event based SIP SUBSCRIBE/NOTIFY framework is used to implement queries and service searches. For clarification purposes, intermediate SIP proxies involved in the messages routing between the different entities are not shown. When a service provider wants to make some information related to itself public, it uses its associated SPA to send a SIP PUBLISH request containing all relevant resource information, which will be included in a document named "Resource Description Document" (RDD). This interaction is shown as phase "a" in Figure 2. The PUBLISH message is sent to the SDA which will store, maintain and aggregate the information provided by different SPAs. This aggregated document that the SDA handles is the "Service Description Document" (SDD).

When an entity (end users or other services) wants to find an SP with specific characteristics, it can subscribe to this information via SIP SUBSCRIBE (phase "b" in Figure 2). The SIP SUBSCRIBE message can contain a body, the "Service Subscription Document" (SSD) which will be used to implement searches, or to restrict the scope of the information the SDS wants to receive. This kind of filtering mechanism is envisaged in the general SIP-Specific Event Notification framework for use with XML. Immediately before a valid subscription is created, the requested information is sent via a SIP NOTIFY message. This information, named the "filtered Service Description Document" (fSDD), typically consists of a subset of the SDD, containing only those RDDs of the resources matching the query. If the subscription document were empty, the whole SDD would be delivered.

Additional to service publications and searches, using this mechanism provides other useful functionalities, as depicted in Figure 2 phase "c": any change in the service provider properties or availability could also be notified during the period of time in which the subscription is valid.

The documents transmitted containing the queries and the service information are in XML format, and the service information is tree-shaped. The queries are formed by using XPath in the current implementation. However, future implementations strive to provide more powerful ontological owl-based syntax, which builds upon XML and is much more expressive than tree-shaped information structures allowed by pure XML.

5 Charging Security and Implementation

To reach an implementable solution for the charging of SAML-based federated VoIP services, the charging approach is based on TICA (Time Interval Calculation Algorithm) [11] and security considerations have also been investigated.

5.1 Charging

TICA enables online charging with important advantages, such as reduction in the number of credit checks and risk-minimization of revenue loss by the service providers, when compared with other approaches, such as hot billing
or DIAMETER Credit-Control Application (CCA). Moreover, it enables support for flexible tariff functions, e.g., a non-linear tariff function depending on several QoS parameters.

The tariff function XML schema specifies the following five top-level elements: PLAN, SERVICE, LOCATION, VALUE and ASSOCIATION. PLAN describes a list of plans available (or formerly available) to customers. SERVICE contains all services offered, e.g., VoIP or web browsing. LOCATION allows location based charging. VALUE contains the information needed to perform mathematical operations with parameters, e.g., duration, bytes transferred, time-of-day and constants to obtain the value to be charged. ASSOCIATION elements associate a PLAN, a SERVICE and a LOCATION to a VALUE.

The main concepts of TICA are adapted to work with the presented SAML payment scheme. In the proposed scenario (cf. Introduction), HSP receives a VoIP call request from S and contacts the FSP, obtaining its tariff. Then, it uses, e.g., a simple addition to calculate the combined tariff\( T = t_{HSP} + t_{FSP} \), where \( t_{HSP} \) is the tariff from the HSP and \( t_{FSP} \) is the tariff from the FSP. Based on \( t_T \), the HSP uses the method as proposed by [11] to estimate resource consumption. To perform resource estimation, the TICA defines time-dependent functions modelling the resource consumption over time. There are three approaches to the resource consumption estimation: TICA 1.0 models a pessimistic estimation by assuming maximum consumption; TICA 1.1 uses statistical prediction based on the last time interval to further reduce credit checks, but allows over-consumption of resources; and TICA 1.2 is similar to TICA 1.1, but uses all preceding intervals to make a more precise estimation at the cost of storing more state information.

After calculating the time interval, the HSP sends to S a request for a token with the desired value for the time interval. It is the task of S to request this token from the PP and to send it to the HSP. The HSP may also authorize the service even if a smaller amount than requested is provided, but, in this case, S must be notified that a shorter time-interval will only be authorized. Once the HSP has received the token, it must issue another token to the FSP to pay for its part of the tariff. When the HSP receives a confirmation from the FSP, it authorizes the user to place the VoIP call. When the value contained in the token is about to expire, the HSP requests a new token from the user. If a new token is not sent, the call is terminated. When the VoIP call is completed, the FSP must send back to the HSP a new token that contains exactly the unused amount. The HSP adds to this value the unused amount it has and returns it to S.

5.2 Security Considerations

The work presented in this paper does not raise any new security issues beyond those related to the SAML-payments that are analyzed in the following subsections. As the security of VoIP systems is undergoing constant investigation and complete issues of journals have been dedicated to VoIP security, e.g., [12], we do not repeat these security discussions. Similarly, we do not address spam [13], as the approach presented in this paper considers only outbound calls that are not affected by spam.

The discovery mechanism does not impose any new security breaches in relation to other security mechanisms that are not subjected to the implementation itself. A wide review of SIP-related security issues can be found in [14].

Regarding SAML payments, not all issues can be discussed due to lack of space, but work on general communications security as well as SAML security [15] has already been carried out. Also, since many security issues are already described in [16], only two general misbehaviours relating to payment assertions are considered here. Thus, regarding the SAML payment processes, there are two major possibilities for misbehaviour: over-spending and data tampering.

5.2.1 SAML Overspending

Overspending could occur when an SP or a consumer attempt to use an assertion more than once. However, this misbehaviour can be avoided if the PP stores the assertion’s ID in local storage. Therefore, after receiving an assertion, whenever a payee requests its validation, the PP defines the
validation request as proof of payment and will mark the assertion as being spent. Thus, any further requests for assertion validation will have the result that this assertion is no longer valid for payment.

5.2.2 SAML Data Tampering

Tampering with the assertion’s data, e.g., enhancing the assertion’s value, can be avoided by establishing a reliable public key infrastructure within the federation. After having generated the assertion, the PP signs it. This signature is checked each time an assertion is used. Thus, any modification of the assertion’s data will be detected and the assertion can then be refused by the payee.

5.3 Implementation

Although the proposed solution integrates known but separated concepts into the the architecture, it’s complexity has resulted in only a partial implementation of this combined approach. Although a full VoIP network has not been deployed, for those SAML-based payments defined, a working prototype has been developed. This prototype is able to interact with several service providers and thus goes further than those SPs just offering VoIP services. This prototype has proven that a reliable federation payment mechanism can be established to support charge-based services.

Furthermore, the TICA approach for charging has been successfully implemented as a prepaid scheme. It shows that the charging technology used (integrated with the Daidalos II [17] architecture) achieves the scheme’s economic incentives. Finally, the Discovery Service implementation carried out in the Akogrimo project [18] was based on the Java JAIN stack for the client side (SPA, SDS). The SDA, also using Java, was built on top of a web server running a proprietary SIP servlet implementation.

6 Conclusion

Within this paper, the concerns for commercial and semi-commercial SPs operating within a SAML-based federation have been described. Based on the fact that the interest of such SPs to join federations will rise, a solution for the payment issue has been found for all those providers not willing to implement a traditional payment application on top of the federation structure, such as semi-commercial service providers or providers offering services requiring micro-payments. A SAML-based payment scheme combines a reliable payment solution with the already established federation technology. Thus, it becomes straightforward to include the payment mechanism into the federation architecture. Moreover, the federation may offer additional advantages, such as discovery services, from which SPs may benefit.

Although described as a combined mechanism, each part can be used standalone. Therefore, Discovery Service, Tariff Function, and SAML Payment can be used separately to support federations in which all aspects may not be required. Thus, although focused on VoIP services, the described elements can also be used for other services.

As an example of commercial or semi-commercial SP services within a federation, VoIP services have been considered. These services have been combined with the benefits of a federation to present the advantages of federation membership for those SPs. It has been shown that SAML-based payment provides an approach, which can be easily deployed to bring payment into federations containing VoIP service providers. However, these kinds of services shall be seen only as an entry point for commercial SPs that wish to offer their other services within a payment-enabled federation.

7 Acknowledgement

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement nr. 215832 (SWIFT). The approach published in this paper was also partly developed within the project DAIDALOS II and the EU IST Network of Excellence EMANICS (IST-2004-NoE-026854).

8 References


[8] Montage Rec and R. En and T Te and De Cette Page: Recommandation G.114 MEAN ONE-WAY PROPAGATION TIME.


