ABSTRACT

The virtual identity concept has been introduced to protect the user’s privacy towards service providers as well as towards access network providers on a service session basis. So far the concept has been only considered from the security and privacy perspective. However, performance analysis is necessary as well to gain insights into the additional costs and their scaling behavior with respect to signaling load and state management. In this paper we propose an analytical model to evaluate the cost introduced by virtual identities. The model allows to quantify the additional signaling overhead and the additional states created by distributing service sessions of users across several virtual identities. We exemplify the use of the model for an authentication, authorization and accounting (AAA) infrastructure based on EAP-TLS [22] and derive figures on the required signaling bandwidth and state overhead to support the virtual identity concept in various service usage scenarios.

Categories and Subject Descriptors


General Terms

Performance, Security

Keywords

Identity management, virtual identities, authentication load, AAA performance

1. INTRODUCTION

Users rely more and more on information and communication technology (ICT) to accomplish their daily life. A multitude of services and applications offered by various service providers (SP) are used for business as well as for private purposes. Normally, SPs require users to create an account for each offered service implying the provision of personal data, like name or email address, to the SP. This results in more and more personal data that is processed and stored by ICT systems and imposes several privacy threats towards the user.

SPs can create detailed user profiles based on the released personal data and based on information that is inherent to today’s Internet-based communication, like the IP address. The IP address can be used to get information about the user’s access network provider and about the location of the user [17]. As long as the same IP address is used for different accounts at the same time, a SP is in the position to link them. In addition it is possible that two SPs collaborate and merge their user profiles in order to derive additional information.

Additionally, access network providers can create very detailed user profiles based on the transported traffic. They are in the position to trace which SPs have been used and can gain additional insights inspecting unencrypted traffic itself. This imposes a severe privacy threat to users, since it requires that access network providers are trustworthy.

The virtual identity (VID) concept [21] is one approach that mitigates the above introduced threats. A user has several VIDs, which can be used for service consumption. Each VID consists only of a subset of all user attributes. In addition, each VID has its own IP address, which prevents that VIDs can be linked based on network layer identifiers. That means, an untrusted access network provider perceives several virtual identities instead of one user with one IP address, which makes the creation of user profiles more difficult. Further improvements towards unlinkability require that the physical layer does not reveal more information, e.g. that the signal strength cannot be used to link different identities.

So far, research on the VID concept focussed mainly on improving the user privacy and security on the network and application layer. However, if the concept is applied in networks and service platforms the performance impact and thus the costs of the VID concept have to be evaluated. In particular the approach to split the user identity into different VIDs is not for free, because it has to be maintained across all layers and across all protocols. For example, each virtual identity must be authenticated and periodically reauthenticated towards the access network provider independent of the others. The same holds with respect to mobility.
management, i.e. mobility updates and handovers have to take place for each VID. Therefore, it is obvious that such an approach creates additional signaling and state management overhead.

This paper examines analytically the additional overhead introduced by the VID concept based on a theoretical model and exemplifies the implications for a federated authentication, authorization, and accounting (AAA) infrastructure using EAP-TLS authentication.

The remainder of this paper is structured as follows. Section 2 introduces related work in the areas of virtual identities and traffic models. Next, Section 3 elaborates the model that is used for the performance evaluation in Section 4. Finally, Section 5 concludes this paper.

2. RELATED WORK

Since this paper examines the performance impact of the virtual identity concept, we first introduce related work on virtual identities in Section 2.1. Afterwards, we present in Section 2.2 an overview on traffic models that characterize the service consumption behavior of individual users.

2.1 Virtual Identities

With the VID concept a user can split his digital identity into several virtual identities. The concept of VIDs is in literature also known as partial identities. [18] defines a partial identity as an identity that reflects a user in a specific role by means of pseudonyms and user attributes linked to it. A partial identity contains only a subset of all user attributes. Thus, partial identities can be designed in way for data minimization and privacy protection on the application layer. To provide unlinkability on the network layer, network anonymization techniques are applied [8].

Our model follows closely the VID concept followed by the EU project Daidalos [21]. In this model the user identity is not only partitioned on the application layer, but also on the network layer. Each VID has its own IP address within a separated network stack [15] running on the user terminal to preserve the privacy of a user on the application layer as well as on the network layer. As a consequence each virtual identity has to be authenticated against the network provider independent of the other VIDs a user has. Therefore, additional signaling load is created on various networking subsystems.

To the best of our knowledge, partial and virtual identity concepts have been solely considered from the security and privacy perspective [13, 9]. The implications on existing communication infrastructures with respect to the additional overhead have not been explored so far.

2.2 Traffic Models

In order to analyze the additional overhead imposed by the VID concept, a user and his service consumption behavior needs to be statistically characterized. In literature various proposals for traffic models, which are also known as work load models, exist that describe the load exposed to systems [10, 19, 16, 6].

From telecommunication research it is known that in simple cases the number of calls per time interval can be described by a Poisson process [11], i.e. the time between two successive calls is negative exponentially distributed. The corresponding call holding time can also be described by a negative exponential distribution [12].

With the advent of the Internet, new services like telnet, electronic mail and file transfer had to be characterized. [7] shows that the interarrival of such service sessions can be modelled by means of negative exponential distributions, whereas the holding time can be modelled by a log-normal distribution.

A lot of research focused on the characterization of HTTP traffic [19, 6]. In [19] it is shown that HTTP sessions can be characterized by means of negative exponential distributions. A generalized approach are hierarchical workload models as proposed in [16]. They differentiate between session layer, function layer, and request layer. In scope of the paper only the session layer is relevant, because it models the different service sessions of users.

The above presented traffic models focus on the traffic observed by system components and cumulate several users. To the best of our knowledge no work exists that provides more details about the individual service consumption behavior of users. In particular for future pervasive systems the users’ behavior cannot be foreseen and gets influenced by the introduction of the VID concept. Therefore, we model in Section 3.3 the interarrival time and the service holding time by means of negative exponential distributions as a first approximation.

3. IDM SYSTEM MODEL

We first provide an overview on the assumed identity management (IdM) system model. Afterwards, the work flow to consume a service is presented in Section 3.2. Finally, we describe in Section 3.3 the service consumption model based on the results from Section 2.2.

3.1 Overview

Figure 1 provides an overview on the assumed identity management system model. A user has a number of VIDs that are provided by an identity provider (IdP) based on a contract. Each virtual identity is represented on the network layer by means of an IP address and a corresponding L2 identifier, e.g. a VID specific MAC address [15]. The access network provider assigns IP addresses to the VIDs based on successful authentication against the IdP. That means for every VID that is in use, a separate authentication procedure has to take place. This is typically achieved by a set of protocols like EAP [3], 802.1X [1] and RADIUS [20] and a federated identity management infrastructure. Figure 1 exemplifies a representative configuration for such an infrastructure that consists of the supplicant in the user terminal, the authenticator residing in the access point or access router and set of interconnected authentication, authentication and authorization (AAA) servers. In the following we refer to all these components as AAA infrastructure.

A user can select each of his VIDs for consuming services provided by various service provider (SP). The SP trusts the IdP regarding successful authentication (Single Sign-On). In consequence, an access network provider as well as a service provider cannot easily link different virtual identities belonging to the same user neither based on network identifiers nor on application layer identifiers and attributes.

The number of VIDs that a user should have is governed by two conflicting factors: privacy versus cost and performance. From a privacy point of view the number of virtual identities should be large in order to have an optimized identity for the given context. This allows for data minimization

46
Figure 1: System model overview

and reduces the probability of linking two different VIDs. In the extreme case a VID is used exactly once for service consumption.

On the other hand the number of VIDs per user has a direct impact on the capital and operational expenditures of the access network provider and of the IDP. If a user can have several identities active in parallel, the authentication and periodic reauthentication signaling load as well as the amount of state information within network components are increased. Therefore, additional infrastructure is needed to support the virtual identity concept.

3.2 Work Flow

Given the user’s intention to use a selected service we can divide the workflow into two phases [4]: the service preparation phase and the service consumption phase. The service preparation phase consists of at least two steps.

First, the user has to select one out of his \( N \) VIDs to consume the service. The selection of VIDs should adhere to privacy considerations based on privacy policies.

Second, the selected VID needs to be potentially activated depending on the current state. Either the VID is already used in other service sessions, i.e. the VID is active and no activation is required, or the VID is not active and needs to be activated. Activation entails to authenticate against the access network and the IDP through additional signaling messages.

If the last service session that is using a specific VID terminates, the VID is deactivated. That means the VID is disassociated from the access network and from the IDP. For future use it needs to be reactivated. Basically, it is possible to keep VIDs active without service sessions. However, this affects among others the interaction with mobility protocols and is subject for future studies.

3.3 Service Consumption Model

The service consumption model shown in Figure 2 details the first phase of the work flow introduced in Section 3.2 for an individual user. The model supports \( m \) different service classes. Each service class \( i \) models a set of services that have independent identical service holding time distributions and their interarrival behaviour follow the same independent interarrival time distribution.

The interarrival time distribution characterizes the time between two service session requests (mean: \( d_i \)), whereas the service holding time distribution describes how long the service is consumed (mean: \( h_i \)). With the assumption from Section 2.2 that interarrival times and service holding times are negative-exponentially distributed, we can characterize the distributions by the mean interarrival rate \( \lambda_i = 1/d_i \) and the mean termination rate \( \mu_i = 1/h_i \).

With the ”Law of the Conservation of Flow” [14] the superposition of all service classes in the service usage multiplexer leads to a negative exponential distributed interarrival time with with a mean interarrival rate of \( \lambda = \sum_{i=0}^{m} \lambda_i \).

In contrast, the superposition of all termination rates results in a more complex distribution. For simplicity we assume as an approximation that \( \mu = 1/\sum_{i=1}^{m} \mu_i h_i \) is negative exponential distributed.

Finally, service session requests have to be split on one of the user’s \( N_{VID} \) VIDs. Due to the lack of statistical data on VID selection and for simplicity, we assume a discrete uniform distribution across all \( N_{VID} \). Based on [14] we obtain that \( \lambda_{VID_i} = \lambda_{VID} \).

Further, we assume that the VID concept has no influence on the service consumption behavior of users. That means the load \( A = \lambda/\mu \) resulting from the superposition of all service classes does not depend on \( N_{VID} \). The load characterizes the mean number of service session that a user simultaneously has.

4. PERFORMANCE EVALUATION

Based on the in Section 3.3 introduced service consumption model, we are in the position to quantify the overhead introduced by the VID concept. We introduce the used key metrics followed by a mathematical model based on Markov chains and apply the results on the in Section 3.1 introduced system model.

4.1 Metrics

We selected three key metrics for the quantification of the performance impact caused by the VID concept:

- **VID activation rate**: The mean VID activation rate \( AR \) quantifies how often VIDs are activated per time unit. It is used to determine the signaling overhead caused by the activation of VIDs. In accordance to \( AR \) we can define a VID deactivation rate. Since we
assume a stationary system, which means that the obtained results are independent of the particular instant of observation, the mean VID deactivation rate has to be the same as $A R$.

- **Authentication load**: The authentication components (Authenticator, AAA server) are only involved during the VID activation. Therefore, the authentication load $A_{Auth}$ is significantly lower than $A$ and can be obtained by multiplying $A R$ with a factor $h_{Auth}$.

- **Mean number of active VIDs**: The mean number of active VIDs $E[N_{act}]$ gives the average number of VIDs that are simultaneously active. $E[N_{act}]$ can be used to quantify the additional state that has to be managed by network components and the signaling for reauthentication.

### 4.2 Mathematical Model

#### 4.2.1 Overview

The above introduced user model can be described in an analytical way by means of infinite n-dimensional Markov chains. The number of required dimensions is $N_{VID}$. Figure 3 illustrates the necessary states and the corresponding transitions for the case of $N_{VID} = 2$.

Each state $X(t) = (x_1, x_2, \ldots, x_{N_{VID}})$ describes the number of active service sessions $x_i$ per VID $i$ at time $t$. The system state transits during an infinitesimal small time interval $dt$ into a state $X(t + dt) = (x_1, x_2, \ldots, x_{N_{VID}})$ with a rate $\lambda_{VID,i} = A/N_{VID,i}$, i.e. a newly initiated service session uses VID $i$. A transition into state $X(t + dt) = (x_1, x_2, \ldots, x_{N_{VID}})$ takes place with rate $x_i \cdot \mu$, if one of the $x_i$ service sessions using VID $i$ terminates.

From the Markov chain model and based on the BCMP theorem $[5]$, we obtain the probability to be in state:

$$p(x_1, x_2, \ldots, x_{N_{VID}}) = p(0, \ldots, 0) \left( \frac{A}{x_1 \cdot x_2 \cdot \ldots \cdot x_{N_{VID}}} \right)$$

(1)

$$p(0, \ldots, 0) = \frac{1}{e^A}$$

(2)

#### 4.2.2 VID Activation Rate

A VID is only activated if it is not used in a prior service session. That means only transitions from $X(t) = (x_1, \ldots, 0, \ldots, x_{N_{VID}})$ to $X(t + dt) = (x_1, \ldots, 1, \ldots, x_{N_{VID}})$ are of interest. In consequence we can group all states in one of $2^{N_{VID}}$ different macro states, in which the same VIDs are active or inactive independent of the number of service sessions per VID. Whether a VID is active is illustrated in Figure 4 by the * symbol for the case of $N_{VID} = 3$, i.e. $x_i > 0$. The edges between the macro states represent VID activation events that are used to calculate $A R$ in combination with the probability to be in one of the macro states.

Due to the symmetry of the macro states, only the number of active VIDs $N_{act}$ per macro state matters for the probability calculation. The probability $p_{k,VID}$ to be in one macro state in which exactly one VID is active can be calculated as:

$$p_{1,VID} = p(\ast, 0, \ldots, 0) = p(0, \ldots, \ast, \ldots, 0)$$

$$= \frac{1}{e^A} \sum_{i=1}^{N_{VID}} \frac{A^i}{i!}$$

$$= \frac{1}{e^A} \left( e^{N_{VID}} - 1 \right)$$

(3)

For the following calculations we have to take into account all macro states in which one VID is active.

$$P(N_{act} = 1) = \left( \frac{N_{VID}}{1} \right) p_{1,VID}$$

(4)

This can be generalized for $k$ VIDs:

$$p_{k,VID} = p(\ast, \ldots, \ast, 0, \ldots, 0) = \frac{1}{e^A} \cdot (e^{N_{VID}} - 1)^k$$

(5)

$$P(N_{act} = k) = \left( \frac{N_{VID}}{k} \right) p_{k,VID}$$

(6)
4.2.3 Authentication Load

(8) indicates that the VID activation rate \( AR \) depends on the load and the mean service termination rate, i.e., the longer the service holding time, the smaller the activation rate. To reflect the actual authentication load \( A_{Auth} \) we introduce \( h_{Auth} = \mu \) and obtain

\[
A_{Auth} = AR \cdot h_{Auth} = \frac{AR}{\mu} = \frac{A}{e^{N_{VID}}} \quad (8)
\]

Figure 5 shows the effect of the number of VIDs on \( A_{Auth} \) in dependence of the load \( A \). We can distinguish two different effects. For low load situations, the probability is high that a new service session triggers the activation of a VID, i.e., \( A_{Auth} \) is high. On the other hand, the higher the load \( A \) the higher the probability that the selected VID is already active, which means that no VID activation takes place.

We obtain a maximum authentication load \( A_{Auth,max} \) on the authentication components for \( A = N_{VID} \)

\[
A_{Auth,max} = N_{VID} \cdot e^{-1} \quad (9)
\]

4.2.4 Mean number of active VIDs

From (5) we can derive the the mean number of active VIDs in dependence of the offered load \( A \).

\[
E[N_{act}] = \sum_{k=0}^{N_{VID}} k \cdot P(N_{act} = k) \quad (10)
\]

\[
= N_{VID} \left( 1 - e^{-\frac{A}{N_{VID}}} \right)
\]

Figure 6 presents the mean number of active VIDs \( E[N_{act}] \) in dependence of the load \( A \). We perceive that for low load values the number of VIDs \( N_{VID} \) has not a tremendous effect, i.e., it does not matter whether a user 3 or 10 VIDs. Low load values mean that the number of simultaneous service sessions is lower than the number of available VIDs.

In contrast, for high load situations (11) converges to \( N_{VID} \), i.e., all VIDs are active. Therefore, the overhead caused by the VID concept is the same as if the number of user is multiplied by \( N_{VID} \).

4.3 Impact on AAA infrastructure

In the following we illustrate the application of this model for the initial architecture depicted in Figure 1 and will focus on the signaling bandwidth needed at the AAA servers. We assume that the processing capacity of the AAA server does not limit the performance. As stated in [2], a RADIUS-based AAA server is capable of several 1000 requests per second.

At first, we introduce the used protocols for the given scenario in Section 4.3.1. Afterwards, more details on the assumed scenario are provided in Section 4.3.2 followed by the presentation of the results in Section 4.3.3.

4.3.1 Used Protocols

We assume that each VID is authenticated against the access network provider by means of EAP-TLS [22], i.e. every VID has an associated certificate. The authenticator exchanges the EAP messages with the AAA server of the access network provider, which in turn forwards the messages to the identity provider as shown in Figure 7. The EAP messages between the authenticator and the AAA server and between the AAA servers are encapsulated in RADIUS [20] messages. Table 1 gives the corresponding message sizes that have been measured in an example scenario. The given values reflect the size of the exchanged IP packets if not stated otherwise.

Given the assumption that the maximum transmission unit of the underlying communication channel is limited (e.g., 1500), larger messages will be fragmented. Message number 4 will be fragmented into three EAP messages, because it contains a certificate chain consisting of three certificates.
Table 1: Assumed message sizes

<table>
<thead>
<tr>
<th>No</th>
<th>Message Type</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>EAP Request Identity</td>
<td>5 Byte(^1)</td>
</tr>
<tr>
<td>1</td>
<td>EAP Response Identity</td>
<td>62 Byte(^2)</td>
</tr>
<tr>
<td>2</td>
<td>EAP Request Start TLS</td>
<td>58 Byte(^2)</td>
</tr>
<tr>
<td>3</td>
<td>Client Hello</td>
<td>155 Byte(^2)</td>
</tr>
<tr>
<td>4</td>
<td>Server Hello</td>
<td>3938 Byte</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(3 ⋅ 1312 Byte(^2))</td>
</tr>
<tr>
<td>5</td>
<td>Client Finish</td>
<td>1022 Byte(^2)</td>
</tr>
<tr>
<td>6</td>
<td>Server Finish</td>
<td>79 Byte(^2)</td>
</tr>
<tr>
<td>7</td>
<td>EAP Response</td>
<td>58 Byte(^2)</td>
</tr>
<tr>
<td>8</td>
<td>EAP Success</td>
<td>52 Bytes(^2)</td>
</tr>
</tbody>
</table>

Table 2: Different User types

<table>
<thead>
<tr>
<th>User type</th>
<th>(N_{VID})</th>
<th>(A) per user</th>
<th>(1/\mu)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordinary User</td>
<td>1</td>
<td>1</td>
<td>1h</td>
</tr>
<tr>
<td>Privacy Aware User</td>
<td>10</td>
<td>1</td>
<td>1h</td>
</tr>
<tr>
<td>Pervasive User</td>
<td>1</td>
<td>10</td>
<td>360s</td>
</tr>
<tr>
<td>Privacy Aware Pervasive User</td>
<td>10</td>
<td>10</td>
<td>360s</td>
</tr>
</tbody>
</table>

that allow to authenticate the A4C server. Each of the EAP messages will be acknowledged, which is not shown in Figure 7, to cope with unreliable transmission channels.

4.3.2 Scenario

In order to get additional knowledge on the consequences caused by the VID concept, we compare four different user behaviors and calculate the imposed bandwidth requirements and the state overhead on the AAA servers.

- **Ordinary User**: An ordinary user has only one VID and uses this VID for all of his service sessions. We assume that the he activates his VID once a hour.
- **Privacy Aware User**: A privacy aware user has the same service consumption behavior as an ordinary user, but distributes his service sessions across 10 VIDs.
- **Pervasive User**: The pervasive user consumes much more services, i.e. higher load, than the ordinary user. He is not privacy aware and has only one VID.
- **Privacy Aware Pervasive User**: The privacy aware pervasive user is the most powerful user. He makes use of all of his 10 VIDs to protect his privacy and consumes a lot of services.

Table 2 provides an overview on the four different user behaviors.

4.3.3 Results

From Table 1 we can derive that one VID activation requires 1297 bytes in downstream direction and 4087 bytes in upstream direction. Since the upstream direction is dominating, we focus on it and use (11) to estimate the imposed bandwidth requirements on the AAA servers for \(N_{User} = 1,000,000\) users connected to the platform.

\[
BW_{up} = AR \cdot N_{User} \cdot 4087\text{byte} \tag{11}
\]

Regarding the additional state overhead, we introduce the total amount of active users \(N_{tot,act}\) in (12).

\[
E_{N_{tot,act}} = N_{User} \cdot E[N_{act}] \tag{12}
\]

Table 3 contains the results for \(N_{User} = 1,000,000\). If an operator supports 10 virtual identities for ordinary users the bandwidth requirements on his AAA servers are increased by a factor of approximately 2.5. The total amount of active users is increased 1.5 times.

In case of pervasive users the requirements on the AAA infrastructure are decreased. Since, the user is almost always authenticated to the platform, the bandwidth requirements are very low. This is also reflected in the total amount of active users.

In particular interesting is the impact of virtual identities for pervasive users, i.e. privacy aware pervasive users. Due to the fact that the service sessions are now distributed on 10 different identities, the bandwidth requirements are increased by several orders of magnitude.

In consequence an identity provider has to know the user traffic models to quantify the impact of virtual identities on his infrastructure as well as on the infrastructure of access network operators.

Regarding the number of states that have to be maintained by network components, we have to cope with 10-times the number of states compared between pervasive users and ordinary users. The privacy-aware user causes only 2.6-times the number of states compared to the ordinary user.

5. CONCLUSION AND OUTLOOK

The introduced mathematical model based on Markov chains allows the quantification of the performance impact caused by the virtual identity concept, which has been targeted to increase the security and privacy of users. We are in the position to determine how often VIDs are activated and how many VIDs are active at the same time. This allows the quantification of the created signaling traffic and supports the dimensioning of network components to cope with additional state caused by the identity partitioning approach. We exemplified the use of the proposed model for an AAA infrastructure and perceived a significant increase of signaling traffic compared to the model without virtual identities.

Even if we are not in the position to forecast the user behavior in the future, we have a measure that allows us to evaluate various scenarios in order to support the dimensioning of the infrastructure. Moreover, the model can serve
as a foundation to answer exceeding questions. For example, the impact of the user mobility on the VID concept is not yet explored and will subject of future work to estimate the consequences on the signaling infrastructure as well as on user’s privacy.

6. ACKNOWLEDGEMENTS

This work was supported in part by the Daidalos project and in part by the SWIFT project. Both projects are funded by the European Union Framework Programme 6 and 7, respectively. The authors thank Andreas Reifert for valuable comments and stimulating discussions.

7. REFERENCES


