Making SIP Make Cents

JASON FISCHL, COUNTERPATH, AND HANNES TSCHOFENIG, SIEMENS

The Session Initiation Protocol (SIP) is used to set up realtime sessions in IP-based networks.¹ These sessions might be for audio, video, or IM communications, or they might be used to relay presence information. SIP service providers are mainly focused on providing a service that copes that provided by the PSTN (public switched telephone network) or the PLMN (public land mobile network) to the Internet-based environment.

Some advantages of using SIP for this purpose are:

- SIP-compliant elements, including hardphones, softphones, PSTN gateways, or proxies can be used on the network.

- SIP service providers can easily peer with each other at very low cost over existing IP network connections.
- IM and presence service providers are able to federate with each other.

Despite these many benefits, however, SIP systems lack an integration with recently introduced identity management solutions in combination with realtime accounting and payments. This article describes how existing SIP deployments address accounting and payments. It also presents a proposal for P2P (peer-to-peer) realtime accounting and payments using SIP in the spirit of federated identity management.

P2P payments using SIP could enable new classes of applications and business models.

Making SIP Make Cents

CHALLENGES

In a traditional e-commerce environment, customers set up direct relationships with merchants. For example, a customer creates an account with a merchant (such as Amazon.com) and provides credit card and user identification information that can be stored for future purchases with the same merchant.

A typical ITSP (Internet telephony service provider) that uses SIP focuses mainly on the delivery of basic services and the features of a telecommunications company, such as voicemail and call waiting. The most common ITSP model is blended with flat-rate monthly pricing, depending on the plan, and per-minute charges for some types of calls. Many ITSPs are moving to a mostly flat-rate plan where all of the minutes are included for an extended geographic region for calls that terminate in the PSTN, and all of the calling features are included with the basic flat rate. International calls that terminate in the PSTN will still have a per-minute charge billed at the end of each month (i.e., a postpaid approach).

Other ITSPs, such as Skype, use a prepaid approach. The customer pays a certain amount of money in advance, then the ITSP rates individual calls in realtime and debits the appropriate amount from the customer's account. To deter fraud, the system must be prepared to cut off a call in realtime if the account balance reaches zero. A call manager might play announcements offering the ability to recharge the account.

Such a system must also be prepared for end users using the same account from different endpoints concurrently and must either disallow this behavior or account for it. Some network operators want to prevent users from directly communicating with each other without involvement of the operator via SIP signaling. This requires additional protocol interaction between SIP entities and routers at the border of the network administrator's domain.

As shown in figure 1, the following are the components of a traditional SIP-based ITSP accounting management system (we use terminology from RFC 2975² as much as possible):

- **SIP server.** This server offers access to the credential database for user authentication, potentially using pro-

tol extensions such as Diameter SIP Application (RFC 4740³) or RADIUS Extension for Digest Authentication (RFC 4590⁴).

- **Call routing engine (typically a SIP proxy).** This entity decides how to route the call based on the dialed address or number.

- **SBC (session border controller, typically a SIP signaling element and a media relay).** This network entity acts as a controller between the ITSP's network and the PSTN gateway. It may also communicate with the RADIUS server to authorize the caller and to account for each call. This element is also responsible for terminating the call in realtime when there are insufficient funds. This entity plays the role of a RADIUS or Diameter client for (real-time) accounting.

- **RADIUS or Diameter Client/Server.** A RADIUS (remote authentication dial-in user service) or Diameter server is responsible for receiving user connection requests, authenticating and authorizing the user, and then returning all configuration and authorization information necessary for the client to deliver service to the user. The RADIUS or Diameter server in turn communicates with the billing and payment system. The role of a RADIUS or Diameter client may be played by an SBC, a SIP proxy, or a SIP server. The RADIUS or Diameter client transmits accounting records at the beginning, during, and at the end of the call. Protocols such as Diameter Credit Control⁵ and RADIUS Prepaid Extensions⁶ may be used to provide realtime accounting capabilities. Non-realtime accounting is offered via the Diameter Base Protocol (RFC 3588⁷) and RADIUS Accounting (RFC 2866⁸).

- **Billing server.** This entity is responsible for the act of preparing an invoice.

- **Payment server.** This entity is responsible for all e-commerce transactions, such as credit card transactions to add funds to a prepaid account or to process the monthly fee in a prepaid system. The term *payment* refers to the transfer of monetary amounts resulting from the accounting process.

There are a number of challenges with the traditional approach:

Most VoIP deployments today focus on simpler pricing models. In these models, calls over the Internet are not charged at all. Users are charged only for calls that terminate in the PSTN. This makes some of the more complicated RADIUS and Diameter usage obsolete.

Realtime accounting adds a lot of protocol complexity to the overall system. Developing a scalable accounting and payment system is challenging because of the realtime and high-availability requirements. In the worst

case, every call requires a realtime interaction with the back-end infrastructure.

Many VoIP provider networks are still islands with a limited amount of inter-provider communication. This requires users either to subscribe to multiple ITSPs or to use the PSTN as a universal interconnection network. The latter prevents advanced features (such as video or presence services) and more sophisticated codecs from being used.

Traditional payment mechanisms have costs associated with each payment transaction. These costs are high enough that micropayments on the order of one cent are not feasible. This eliminates an entire class of services that could be offered to customers.

By way of comparison, in the HTTP-based environment, Web-based payment schemes such as PayPal and Google Checkout have provided merchants with a means of accepting P2P payments through the Web without requiring customers to have direct relationships with the merchants. Merchants that want to use these payment services need to implement their Web storefronts using a proprietary API (application programming interface) provided by the payment service provider. These types of payment services work well when the seller is selling goods or services from a Web page. It is possible to use such a payment mechanism to provide a SIP-enabled service, such as a phone call, but it is challenging because of the realtime nature of the transaction. It would also be practically limited to paying the ITSP's monthly flat-

rate fee. Demanding an interaction with the back-end infrastructure with every single call is also not desirable. Furthermore, there is no standard interface or protocol defined for making payments. Proprietary APIs tend to lock clients into specific payment service providers.

REQUIREMENTS

In July 2004, Cullen Jennings from Cisco Systems and Gyuchang Jun from Bitpass, a micropayment provider startup company, proposed an initial Internet draft for a simple mechanism to address some of the limitations imposed by more traditional approaches.⁹ At roughly the same time, the PayCircle organization investigated the applicability of Liberty Alliance protocols for payment.¹⁰

In their SIP payment Internet draft, Jennings and Jun addressed the following requirements:

The approach should be optimized for low-value transactions where the consequences of a cheating merchant are minimal to the customer. The customer can decide not to purchase services from the merchant again with a minimal monetary loss. This assumption leads to much simpler implementations.

The customer and merchant should not need to have a direct trusted relationship with each other. Furthermore, customers should be permitted to remain completely anonymous to the merchants. They should be able to use a trusted third-party payment service provider as a clearinghouse. A one-time purchaser of an inexpensive service need not trust the merchant. At most, the purchaser runs the risk of losing an initial small payment to the merchant.

A merchant should be able to support multiple payment providers. This choice of payment providers should be communicated to customers through a standardized protocol. By doing this, the merchant appeals to a wider range of customers. For example, a merchant selling language-training service from the United States might use a U.S.-based payment provider when offering services to U.S.-based customers but might want to use a Chinese payment provider

Making SIP Make Cents

when selling the same service to Chinese customers.

The customer needs to be able to find out the price before making a purchase. This is a form of advice-of-charge but offers a stronger enforcement from the customer side. For example, Alice places a call to a merchant offering realtime video horoscopes. The horoscope vendor offers Alice the option of purchasing one basic horoscope for 50 cents or a deluxe horoscope for \$2. Since Alice has never purchased anything from this vendor, she might decide to try the basic option first. Even if the merchant turns out to be a fraud, Alice is risking only 50 cents. The merchant might even offer Alice her money back if she is not satisfied with the horoscope.

The merchant should be able to support flat-rate, time, volume, and session-based charging. Furthermore, different currencies should be supported. For example, a PSTN termination merchant might offer to terminate phone calls anywhere in South Africa under a number of different models including \$1 USD for up to a one-hour phone call or six cents per minute. These different rate plans could be offered using an American payment provider in U.S. dollars and a German payment provider in Euros. It is also possible to charge in pseudo-currencies, such as those earned in airline loyalty programs. For example, United Airlines could offer a call anywhere in the United States for 1,000 air miles.

The merchant should be able to offer simple refunds. Although another requirement is to be optimized for low-value transactions, it is still necessary to provide refunds—for example, when a customer agrees to pay for a call to South Africa but the dialed destination is busy. Similarly, the customer may have prepaid for 20 minutes of per-minute usage but if the call ends prematurely, the merchant may wish to refund some of the payment.

A STANDARDS-BASED SOLUTION

The efforts made by Jennings, Jun, and others at the IETF (Internet Engineering Task Force) have outlined the requirements and a proposed solution for a simple SIP-based P2P payment mechanism. The proposed solution is lightweight and intended to be easy for customers, merchants, and payment providers to implement.¹¹ First, some definitions:

Customer. The buyer of the service. A SIP softphone or hardphone could take on this role. This is typically a SIP UAC (user-agent client), although a SIP proxy in the same administrative domain as the customer can perform this function on behalf of the customer.

Merchant. The seller of the service, which is the entity wishing to be paid. A PSTN gateway could take on this role. This is typically a SIP UAS (user-agent server), although a SIP proxy in the same administrative domain as the merchant can perform this function on behalf of the merchant.

Payment provider. A third party that handles the transfer of funds from customer to merchant and vice versa. The customer communicates with the payment provider using HTTPS.

Payment offer. The information sent from the merchant to the customer describing what payment is required, including rates, supported currencies, and trusted payment providers.

Payment request. The information sent from the customer to the payment provider requesting a payment from the customer's account to the merchant's account.

Receipt. A digitally signed document sent from the payment provider to the customer in response to a request for payment, and passed on to the merchant. The receipt tells the merchant that the customer has successfully made a payment. The payment provider needs to digitally sign the receipt to allow the merchant to detect unauthorized modifications. The current proposal is to use a SAML (Security Assertion Markup Language) assertion for the receipt. SAML is an XML standard for exchanging authentication and authorization data between security domains.

Refund. The merchant can use a payment request to transfer funds back to the customer.

The basic message flow is described in figure 2. The signaling between the customer and the merchant is done using simple extensions to the SIP protocol as described in Jennings and Jun's SIP payment Internet draft. The signaling between the customer and the payment provider is done using HTTPS. The digitally signed receipt, which is returned in the HTTPS response from the payment provider, is a SAML assertion. This receipt is then provided in a subsequent SIP request back to the merchant. The merchant then verifies the receipt and allows service. If the customer paid only a fraction for the total service usage, then it is up to the customer to make another payment before the merchant terminates the session. This extension is accomplished by repeating the same set of steps as in the original payment.

Note that the protocol interaction shown in figure 2 does not require an interaction between the customer and the payment provider for every customer-to-merchant SIP interaction. A receipt from the payment provider may have a validity that spawns many SIP sessions. Furthermore, it is worth noting that figure 2 shows a push approach only. A pull-based approach requires an interaction between the merchant and the payment provider. This can be accomplished by using references to assertions, but a detailed description is outside the scope of this article (see the Internet draft¹² for more details).

APPLICATIONS

In addressing some of the fundamental limitations with traditional realtime accounting and payment processing, the approach described here could enable a wide array of new applications. For example, having a caller place a “payment at risk” might help solve the spam problem in VoIP systems. One of the key reasons that spam e-mail has proliferated has been the almost zero cost associated with sending it. Spam over Internet telephony and instant messaging (spit and spim) is anticipated to be an even more significant problem. Because of the realtime nature of the interrupt caused by an unsolicited phone call or instant message, it is likely to be an even more annoying problem to end users.

One factor that exacerbates the problem is the ability to create new accounts (and therefore identities) for little or no cost with one of these ITSPs. Consequently, the solution to the problem needs to assume that new, previously unknown identities cannot be trusted by default. Hence, the usefulness of blacklists is limited.

Increasing the cost to the spammer can mitigate this problem. The first time that a user receives an unsolicited call or instant message, the recipient can send back a payment request for a small fee—maybe five cents—to the caller. The payment could even be made to the recipient’s favorite charity. Once the payment has been made, the recipient can now add the caller’s identity to a whitelist for all future communication. Furthermore, the recipient could refund the payment to the caller. The payment service provider that brokered the transaction would receive a fee for both transactions in this case. Note that the identity of the caller must be difficult to forge or impersonate.

Another interesting application of this approach creates a P2P market for realtime communication services. For example, language-training services could be provided on a SIP network between parties that have no direct commercial relationship. An English language student could be looking for an English language teacher and be willing to pay for this service. The seller of the training service could choose among a variety of models for charging for the service: by the minute, by the hour, or for a flat fee. The conversation could include audio, video, and/or IM.

To find a particular seller, the buyer would use a directory that includes rating information similar to the type of ratings that eBay provides for its buyers and sellers to assist the buyer in choosing a provider. When the buyer first contacts the seller, the seller sends back a request for payment indicating the supported billing models, currencies, and payment service providers. At this stage, buyers can choose which charging model they want to go

with and which payment provider they wish to use. The buyer can even decide to pay for a few minutes of service initially to see how good the seller is and then purchase additional time after the session has started.

After the buyer makes an initial payment, the payment service provider returns a digitally signed receipt to the buyer, which is then sent to the seller. The seller verifies the receipt and the session starts. It is up to the

Making SIP Make Cents

seller to decide the business model. The payment service provider merely acts as a third-party broker for the payment. This type of payment mechanism is optimized for low-value transactions. If the seller does not provide the advertised service, the buyer has lost only a small amount of money and can provide negative feedback on the seller.

ANALYSIS

Diameter and RADIUS extensions^{13,14,15,16} are building blocks being developed within the IETF that offer authentication, authorization, and (realtime) accounting. These generic mechanisms are available not only for network access but also for application services, such as SIP sessions. The message communication pattern that is being followed by RADIUS and Diameter focuses mostly on the interaction between the merchant and the payment provider whereby the customer is not involved except for authentication. With the introduction of a new generation of identity management solutions, such as SAML, the customer can be more involved by routing messages through the customer's device.

New communication patterns with different characteristics need to be investigated. The Jennings and Jun draft presents a proposal that aims to combine the Web-based SAML-based identity management with the capability to perform realtime accounting in SIP supporting a variety of payment mechanisms. The outcome is a pay-as-you-go scheme with a built-in advice of charge giving the end user more control (including privacy) and visibility of authorization decisions. Finally, functionality and complexity are pushed to the end host, and we believe that the network infrastructure is more robust and more scalable as a consequence.

Even though a paradigm change as envisioned by the identity management community behind SAML (and similar proposals) seems to hold great promise for telephony applications and beyond, some questions remain. Among these: Is it feasible for a payment provider to offer a micropayment solution in a cost-effective way? People have been talking about micropayments for many years, but no micropayment company seems to be taking off.

In January at Microsoft's "Think Week," Bill Gates announced a development plan for a new payment

system that addresses many of the challenges outlined earlier and that "will be cheaper than credit card transactions, making it possible for companies to charge small fees for Web-based content and services they now offer for free," according to Dow Jones Newswire reports. SAML-based identity management solutions are being used more widely (beyond the enterprise and business-to-business deployment environment) and are evolving from asserting identities to authorization and attribute statements. These systems will also have to provide integrated accounting and payment functionality in the future. Combining a cheap micropayment solution with an evolved identity management system could prove to carry weight in the real world. Interesting applications and business models can be enabled with a payment model applicable to Web- and SIP-based services.

Another unresolved issue concerns fraud. With some of the simplifying assumptions that are made, there is potential for customers to lose money to fraudulent merchants. Since the customer does not have a direct, trusted relationship with the merchant, the amount of money at risk here is constrained and is likely an acceptable risk to the customer. The consequences for a cheating merchant are a poor rating in public directories.

The refund mechanism is also very simple and does not allow the merchant or customer much flexibility. Again, by optimizing for a small transaction, a complex refund mechanism is not required. This approach tends to encourage different business models than are traditional in telecommunications providers, but as many are predicting, basic voice will be free before too much longer. New business models are needed for telecommunications companies to survive.

Another question that comes to mind is how the SIP infrastructure providers can make money in this model. In the examples described here, the merchant is not always the same company as that providing basic SIP service to the customer. For example, Alice is registered with a SIP ITSP called example.com, which can offer basic termination using the mechanism described here but may also want to provide interconnection over the Internet to a variety of third-party merchants. Example.com may choose to charge a fee to the merchants for routing the call to them since it has access to the payment offer in the SIP signaling. The means of charging the merchants is out of the scope of this article, but it is easy to see an interesting business model for the ITSP that does not involve providing the actual service but, instead, facilitating it. The key services that the ITSP offers here are identity and routing. The identity service implemented by the

ITSP integrity protects the receipt that is included in the request relayed to the merchant. If the merchant decides that the ITSP cannot be trusted, then it can blacklist all requests coming from the ITSP's domain.

CONCLUSION

New payment models that leverage growing a widely deployed SIP infrastructure hold much promise for the future of communications, enabling service providers, merchants, and users to benefit from increased flexibility and efficiency. To be fully integrated into existing systems and into the fabric of our lives, however, micropayment service providers will need to proliferate, and the technical and business barriers to entry will have to be reduced. As ongoing work in this and other areas of SIP integration proves to yield promising results, supporters will continue to see increased interest within various sectors of the industry. Q

ACKNOWLEDGMENTS

The authors would like to thank Cullen Jennings and Gyuchang Jun for their work on the SIP payment Internet draft, Jeff Hodges for his help with SAML, the IETF SIP-PING working group for their feedback, and Bertolt Eicke, Karsten Luetttge, Thomas Hickethier, and Fatih Eyüp Nar for their review comments. We would also like to thank Layya Halawi, Matthias Gsottberger, and Marcos Dytz for their work on a prototype implementation of the approach outlined in the SIP payment Internet draft.

REFERENCES

1. Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., Schooler, E. 2002. SIP: Session Initiation Protocol. RFC 3261 (June).
2. Aboba, B., Arkko, J., Harrington, D. 2000. Introduction to accounting management. RFC 2975 (October).
3. Garcia-Martin, M., Belinchon, M., Pallares-Lopez, M., Canales-Valenzuela, C., Tammi, K. 2006. Diameter Session Initiation Protocol (SIP) Application. RFC 4740 (November).
4. Stermann, B., Sadolevsky, D., Schwartz, D., Williams, D., Beck, W. 2006. RADIUS extension for digest authentication. RFC 4590 (July).
5. Hakala, H., Mattila, L., Koskinen, J-P., Stura, M., Loughney, J. 2005. Diameter credit-control application. RFC 4006 (August).
6. Lior, A., Yegani, P., Chowdhury, K., Tschofenig, H., Pashalidis, A. 2006. Prepaid extensions to remote authentication dial-in user service (RADIUS), draft-lior-radius-prepaid-extensions-11.txt (work in progress, June).
7. Calhoun, P., Loughney, J., Guttman, E., Zorn, G., Arkko, J. 2003. Diameter Base Protocol. RFC 3588 (September).
8. Rigney, C. 2000. RADIUS accounting. RFC 2866 (June).
9. Jennings, C., Fischl, J., Tschofenig, H., Jun, G. 2006. Payment for services in Session Initiation Protocol (SIP). draft-jennings-sipping-pay-05.txt (October).
10. Foll, F., Eicke, B., Luetttge, K. PayCircle white paper on Liberty integration, PayCircle (September); http://www.paycircle.org/downloads/file.php?id=28&kat_id=1
11. See reference 9.
12. See reference 9.
13. See reference 3.
14. See reference 4.
15. See reference 5.
16. See reference 6.

LOVE IT, HATE IT? LET US KNOW

feedback@acmqueue.com or www.acmqueue.com/forums

JASON FISCHL is the CTO of CounterPath Solutions (formerly Xten Networks), a leading provider of SIP softphones. Before joining CounterPath, he was co-founder and CTO of Tel Tel and PurpleComm Inc., a SIP service provider offering residential VoIP, presence, and IM services to users around the world. He was also a founder and CTO of Cathay Networks, which developed a SIP-based IP Centrex system. Fischl is one of the founding board members of both SIPfoundry and Vovida.org. He helped found the reSIProcate open source SIP stack project in 2002 and has been one of its key contributors. He is an active member at the IETF and is working on a number of Internet drafts in the areas of SIP security and payment mechanisms.

HANNES TSCHOFENIG works as a senior research scientist in Siemens Networks GmbH & Co KG, focusing on security and network protocols. He received his degree in applied computer science from the University of Klagenfurt, Austria. He is active in the standardization arena, especially in the IETF where he co-chairs the ECRIT (Emergency Context Resolution with Internet Technologies), the KeyProv (Provisioning of Symmetric Keys), and the IETF DIME (Diameter Maintenance and Extensions) working groups. He is also the secretary of the IETF NSIS (Next Steps in Signaling) working group and a member of the Transport Area Directorate. Tschofenig has co-authored several publications for international conferences, a number of RFCs, and many IETF drafts. © 2007 ACM 1542-7730/07/0300 \$5.00