Network Working Group J. Peterson
Internet-Draft NeuStar, Inc.
Intended status: Informational September 04, 2013
Expires: March 08, 2014

Secure Telephone Identity Threat Model
draft-peterson-stir-threats-00.txt

Abstract

As the Internet and the telephone network have become increasingly interconnected and interdependent, attackers can impersonate or obscure calling party numbers when orchestrating bulk commercial calling schemes, hacking voicemail boxes or even circumventing multi-factor authentication systems trusted by banks. This document analyzes threats in the resulting system, enumerating actors, reviewing the powers available to and used by attackers, and describing scenarios in which those powers are exercised.

Table of Contents

1. Introduction and Scope

As is discussed in the STIR problem statemnt [9], the primary enabler of robocalling, vishing and related attacks is the capability to impersonate a calling party number. The most stark example of these attacks are cases where automated callees on the PSTN rely on the calling number as a security measure, for example to access a voicemail system. Robocallers use impersonation as a means of obscuring identity; while robocallers can, in the ordinary PSTN, block (that is, withhold) their caller identity, callees are less likely to pick up calls from blocked identities, and therefore calling from some number, any number, is prefereable. Robocallers however prefer not to call from a number that can trace back to the robocaller, and therefore they impersonate numbers that are not assigned to them.

The scope of impersonation in this threat model pertains solely to the rendering of a calling telephone number to an end user or automaton at the time of call set-up. The primary attack vector is therefore one where the attacker contrives for the calling telephone number in signaling to be a particular chosen number, one that the attacker does not have the authority to call from, in order for that number to be rendered on the terminating side. The threat model assumes that this attack simply cannot be prevented: there is no way to stop the attacker from creating calls that contain attacker-chosen calling telephone numbers in their signaling. The solution space

therefore focuses on ways that terminating or intermediary elements might differentiate authorized from unauthorized calling party numbers, in order that policies, human or automatic, might act on that information.

Rendering an authenticated calling party number during call set-up time does not entail anything about the entity or entities that will send and receive media during the call itself. In call paths with intermediaries and gateways as described below, there may be no way to provide any assurance in the signaling about participants in the media. In those end-to-end IP environments where such an assurance is possible, it is highly desirable, but in the threat model considered in this document, the threat of impersonation does not extend to impersonating an authorized listener after a call has been completed. Attackers that could impersonate an authorized listener require powers that robocallers and voicemail hackers are unlikely to possess, and historically such attacks have not played a role in enabling robocalling or related problems.

In protocols like SIP, call signaling can be renegotiated after the call has been completed, and through various transfer mechanisms common in telephone systems, callees can easily be connected to, or conferenced in with, telephone numbers other than the original calling number once a call has been set up. These post-setup changes to the call are outside the scope of impersonation considered in this model. Furthermore, impersonating a reached number to the originator of a call is outside the scope of this threat model.

In much of the PSTN, there exists a supplemental service that translates calling party numbers into regular names, including the proper names of people and businesses, for rendering to the called user. These services (frequently termed 'Caller ID') provide a further attack surface for impersonation. The threat model explored in this document focuses only on the calling party number, though presenting a forged calling party number can let the attacker cause a forged 'Caller ID' name to be rendered to the user as well. Providing a verifiable calling party number therefore does improve the security of Caller ID systems, but this threat model does not consider attacks specific to Caller ID, such as attacks on the databases consulted by the terminating side of a call to provide Caller ID, or impersonators choosing to forge a particular calling party number in order to present a misleading Caller ID to the user.

Finally, the scope of impersonation in this threat model does not consider simple anonymity as a threat. The ability to place anonymous calls has always been a feature of the PSTN, and users of the PSTN today have the capability to reject anonymous calls should they wish to.

2. Actors

2.1. Endpoints

There are two main categories of end-user terminals, a dumb device (such as a 'black phone') or a smart device:

Dumb devices comprise a simple dial pad, handset and ringer, optionally accompanied by a display that can show only a limited number of characters (typically, enough for a telephone number and an accompanying name, sometimes less). These devices are controlled by service providers in the network.

Smart devices are general purpose computers with some degree of programmability and the capacity to access the Internet, along with a rich display. This includes smart phones, telephone applications on desktop and laptop computers, IP private branch exchanges, and so on.

There are also various hybrid devices, such as terminal adapters which attach dumb devices to a VoIP service, but which may in turn use auxiliary screens as displays for rich information (for example, some cable deployments use the television screen to render caller ID). These devices expose little programmability to end users.

There is a further category of automated terminals without an end user. These include systems like voicemail services that consume the calling party number without rendering it to a human. Though the capability of voicemail services varies widely, many today have Internet access and advanced application interfaces (to render 'visual voicemail,' to automatically transcribe voicemail to email, and so on).

2.2. Intermediaries

We assume that a call between two endpoints traverses a call path. The length of the call path can vary considerably: it is possible in VoIP deployments for two endpoint entities to send traffic to one another directly, but more commonly several intermediaries exist in a VoIP call path. One or more gateways may also appear on a call path.

Intermediaries forward call signaling to the next entity in the path. These intermediaries may also modify the signaling in order to improve interoperability, to enable proper network-layer media connections, or to enforce operator policy. This threat model assumes there are no restrictions on the modifications to signaling that an intermediary can introduce.

Gateways translate call signaling from one protocol into another. In the process, they tend to consume any signaling specific to the original protocol (elements like transaction-matching identifiers) and may need to transcode or otherwise alter identifiers as they are rendered in the destination protocol.

This threat model assumes that intermediaries and gateways can forward and retarget calls as necessary, which can result in a call terminating at a place the originator did not expect, and that this is an ordinary condition in call routing. This is significant to the solution space, however, because it limits the ability of the originator to anticipate what the telephone number of the respondent will be.

Furthermore, we assume that some intermediaries or gateways may, due to their capabilities or policies, discard calling party number information, as a whole or in part. Today, many IP-PSTN gateways simply ignore any information available about the caller in the IP leg of the call, and allow the telephone number of the PRI line that the gateway happens to use to be sent as the calling party number for the PSTN leg of the call. A call might also gateway to a multifrequency network where only a limit number of digits of automatic numbering identification (ANI) data are signaled, for example. Some protocols may render telephone numbers in a way that makes it impossible for a terminating side to parse or canonicalize a number. In these cases, providing authenticated identity may be impossible. This is not however indicative of an attack or other security failure.

2.3. Attackers

We assume that an attacker has the following powers:

The attacker can create telephone calls at will, originating them on either the PSTN or over IP, and can supply an arbitrary calling party number.

The attacker can capture and replay signaling previously received. [TBD: should this include a passive attacker that can capture signaling that isn't directly sent to it? Not a factor for robocalling, but perhaps for voicemail hacking, say.]

The attacker has access to the Internet, and thus the ability to inject arbitrary traffic over the Internet, to access public directories, and so on.

There are many potential threats in which an attacker compromises intermediaries in the call path, or captures credentials that allow

the attacker to impersonate a target. Those system-level threats are not considered in this threat model, though secure design of systems to prevent these sorts of attacks is necessary for any of these countermeasures to work.

This threat model also does not consider a case in which the operators of intermediaries or gateways are themselves adversaries who intentionally suppress identity or send falsified identity with their own credentials.

3. Attacks

3.1. Voicemail Hacking via Impersonation

A voicemail service allows users calling from their mobile phones access to their voicemail boxes on the basis of the calling party number. An attacker wants to access the voicemail of a particular target. The attacker therefore impersonates the calling party number using one of the scenarios described below.

In all cases, the countermeasure to this threat is for the voicemail service to have an expectation that calls to its service will supply an authenticated identity, and in the absence of that identity, for it to adopt a different policy (perhaps requiring a shared secret to be dialed as a PIN). Authenticated identity alone provides a positive confirmation only when an identity is claimed legitimately; the absence of authenticated identity here is not evidence of malice, just of uncertainty.

If the voicemail service could know ahead of time that it should always expect authenticated identity from a particular number, that would enable the voicemail service to adopt different policies for handling a request without authenticated identity. Since users contact a voicemail service repeatedly, this is something that a voicemail server could learn, for example, the first time that a user contacts it. Alternatively, it could access a directory of some kind that informs verifiers that they should expect identity from particular numbers.

3.2. Unsolicited Commercial Calling from Impersonated Numbers

The unsolicited commercial calling, or for short robocalling, threat is similar to the voicemail threat, except in so far as the robocaller does not need to impersonate any specific number, merely a plausible number. A robocaller may impersonate a number that is not a valid number (for example, in the United States, a number beginning with 0), or an unassigned number. The robocaller may change numbers every time a new call is placed, even selecting numbers randomly.

The countermeasures to robocalling are similar to the voicemail example, but there are significant differences. One important potential countermeasure is simply to verify that the calling party number is in fact valid and assigned. Unlike voicemail services, end users typically have never been contacted by the number used by a robocaller before, so they can't rely on past association to know whether or not the calling party number should always supply authenticated identity. If there were a directory that could inform the terminating side of that fact, however, that would help in the robocalling case.

When alerting a human is involved, the time frame for executing these countermeasures is necessarily limited. Ideally, a user would not be alerted that a call has been received until any necessary identity checks have been performed. This could however result in inordinate post-dial delay from the perspective of legitimate callers. Cryptographic operations and network operations must be minimized for these countermeasures to be practical.

The eventual effect of these countermeasures would be to force robocallers to either block their caller identity, in which case end users could opt not to receive their calls, or to use authenticated identity for numbers traceable to them, which would then allow for other forms of redress.

3.3. Attack Scenarios

Impersonation, IP-PSTN

An attacker on the Internet uses a commercial WebRTC service to send a call to the PSTN with a chosen calling party number. The service contacts an Internet-to-PSTN gateway, which inserts the attacker's chosen calling party number into the CPN field of an IAM. When the IAM reaches the endpoint terminal, the terminal renders the attacker's chosen calling party number as the calling identity.

Countermeasure: out-of-band authenticated identity

Impersonation, PSTN-PSTN

An attacker with a traditional PBX (connected to the PSTN through an ISDN PRI) sends a Q.931 SETUP request with a chosen calling party number which a service provider inserts into the corresponding SS7 CPN field of an IAM. When the IAM reaches the endpoint terminal, the terminal renders the attacker's chosen calling party number as the calling identity.

Countermeasure: out-of-band authenticated identity

Impersonation, IP-IP

An attacker with an IP phone sends a SIP request to an IP-enabled voicemail service. The attacker puts a chosen calling party number into the From header field value of the INVITE. When the INVITE reaches the endpoint terminal, the terminal renders the attacker's chosen calling party number as the calling identity.

Countermeasure: in-band authenticated identity

Impersonation, IP-PSTN-IP

An attacker with an IP phone sends a SIP request to the telephone number of a voicemail service, perhaps without even knowing that the voicemail service is IP-based. The attacker puts a chosen calling party number into the From header field value of the INVITE. The attacker's INVITE reaches an Internet-to-PSTN gateway, which inserts the attacker's chosen calling party number into the CPN field of an IAM. That IAM then traverses the PSTN until (perhaps after a call forwarding) it reaches another gateway, this time back to the IP realm, to an H.323 network. The PSTN-IP gateway puts takes the calling party number in the IAM CPN field and puts it into the SETUP request. When the SETUP reaches the endpoint terminal, the terminal renders the attacker's chosen calling party number as the calling identity.

Countermeasure: out-of-band authenticated identity

3.4. Solution-Specific Attacks

[TBD: This is just forward-looking notes]

Threats Against In-band

Token replay

Removal of in-band signaling features

Threats Against Out-of-Band

Provisioning Gargbage CPRs

Data Mining

Threats Against Either Approach

Attack on directories/services that say whether you should expect authenticated identity or not

Canonicalization attack

4. Acknowledgments

Henning Schulzrinne, Hannes Tschofenig, Cullen Jennings and Eric Rescorla provided key input to the discussions leading to this document.

5. IANA Considerations

This memo includes no request to IANA.

6. Security Considerations

This document provides a threat model and is thus entirely about security.

7. Informative References

[1] Peterson, J. and C. Jennings, "Enhancements for Authenticated Identity Management in the Session Initiation Protocol (SIP)", RFC 4474, August 2006.

[2] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.

[3] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.

[4] Jennings, C., Peterson, J., and M. Watson, "Private Extensions to the Session Initiation Protocol (SIP) for Asserted Identity within Trusted Networks", RFC 3325, November 2002.

[5] Hoffman, P. and J. Schlyter, "The DNS-Based Authentication of Named Entities (DANE) Transport Layer Security (TLS) Protocol: TLSA", RFC 6698, August 2012.

[6] Elwell, J., "Connected Identity in the Session Initiation Protocol (SIP)", RFC 4916, June 2007.

[7] Schulzrinne, H., "The tel URI for Telephone Numbers", RFC 3966, December 2004.

[8] Cooper, A., Tschofenig, H., Peterson, J., and B. Aboba, "Secure Call Origin Identification", draft-cooper-iab-secure-origin-00 (work in progress), November 2012.

[9] Peterson, J., Schulzrinne, H., and H. Tschofenig, "Secure Origin Identification: Problem Statement, Threat Model, Requirements, and Roadmap", draft-peterson-secure-origin-ps-01 (work in progress), July 2013.

[10] Peterson, J., "Retargeting and Security in SIP: A Framework and Requirements", draft-peterson-sipping-retarget-00 (work in progress), February 2005.

[11] Rosenberg, J., "Concerns around the Applicability of RFC 4474", draft-rosenberg-sip-rfc4474-concerns-00 (work in progress), February 2008.

[12] Kaplan, H. and V. Pascual, "Loop Detection Mechanisms for Session Initiation Protocol (SIP) Back-to- Back User Agents (B2BUAs)", draft-ietf-straw-b2bua-loop-detection-01 (work in progress), August 2013.

[13] Barnes, M., Jennings, C., Rosenberg, J., and M. Petit-Huguenin, "Verification Involving PSTN Reachability: Requirements and Architecture Overview", draft-jennings-vipr-overview-04 (work in progress), February 2013.

[14] Rosenberg, J. and H. Schulzrinne, "Session Initiation Protocol (SIP): Locating SIP Servers", RFC 3263, June 2002.

Author's Address

Jon Peterson
NeuStar, Inc.
1800 Sutter St Suite 570
Concord, CA 94520
US

Email: jon.peterson@neustar.biz