Network Working Group
Internet-Draft
Intended status: Informational
Expires: August 9, 2014

J. Peterson
NeuStar, Inc.

February 5, 2014

Secure Telephone Identity Threat Model
draft-ietf-stir-threats-01.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 capabilities available to and used by attackers, and describing scenarios in which attacks are launched.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on August 9, 2014.

Copyright Notice

Table of Contents

1. Introduction and Scope 2
2. Actors 4
2.1. Endpoints 4
2.2. Intermediaries 4
2.3. Attackers 5
3. Attacks 6
3.1. Voicemail Hacking via Impersonation 6
3.2. Unsolicited Commercial Calling from Impersonated Numbers 7
3.3. Telephony Denial-of-Service Attacks 8
4. Attack Scenarios 9
4.1. Solution-Specific Attacks 10
5. Acknowledgments 10
6. IANA Considerations 11
7. Security Considerations 11
8. Informative References 11
Author's Address 11

1. Introduction and Scope

As is discussed in the STIR problem statement [2], the primary enabler of robocalling, vishing, swatting and related attacks is the capability to impersonate a calling party number. The starkest 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 preferable. 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 a callee (human 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 specific number. In this attack, the number is one that the attacker is not authorized to use (as a caller), but gives in order for that number to be consumed or 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. 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.

Securing an authenticated calling party number at 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 of a call. In those end-to-end IP environments where such an assurance is possible, it is highly desirable. However, in the threat model described in this document, "impersonation" does not consider impersonating an authorized listener after a call has been established, such as a third party attempting to eavesdrop on a conversation. Attackers that could impersonate an authorized listener require capabilities 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 SIP and even many traditional telephone protocols, call signaling can be renegotiated after the call has been established. Using various transfer mechanisms common in telephone systems, a callee can easily be connected to, or conferenced in with, telephone numbers other than the original calling number once a call has been established. 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 described

in this document addresses only the calling party number, even though presenting a forged calling party number may cause a chosen 'Caller ID' name to be rendered to the user as well. Providing a verifiable calling party number therefore improves the security of Caller ID systems, but this threat model does not consider attacks specific to Caller ID. Such attacks may be carried out against the databases consulted by the terminating side of a call to provide Caller ID, or by impersonators forging a particular calling party number in order to present a misleading Caller ID to the user.

2. Actors

2.1. Endpoints

There are two main categories of end-user terminals relevant to this discussion, 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 render a limited number of characters (typically, enough for a telephone number and an accompanying name, sometimes less). Although users interface with these devices, the intelligence that drives them lives in the service provider network.
>
> Smart devices are general purpose computers with some degree of programmability, and with the capacity to access the Internet and to render text, audio and/or images. This category includes smart phones, telephone applications on desktop and laptop computers, IP private branch exchanges, and so on.

There is a further category of automated terminals without an end user. These include systems like voicemail services, which may provide a different set of services to a caller based solely on the calling party's number, for example granting the mailbox owner access to a menu while giving other callers only the ability to leave a message. 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

The endpoints of a traditional telephone call connect through numerous intermediary switches in the network. The set of intermediary devices traversed during call setup between two endpoints is referred to as 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 (which is consistent with the observed behavior of such devices).
>
> Gateways translate call signaling from one protocol into another. In the process, they tend to consume any signaling specific of 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; this is an common condition in call routing. This is significant to the solution space, because it limits the ability of the originator to anticipate what the telephone number of the respondent will be (for more on the "unanticipated respondent" problem, see [3]).

Furthermore, we assume that some intermediaries or gateways may, due to their capabilities or policies, discard calling party number information, in whole or 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 used by the

gateway 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 limited 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, but this is not indicative of an attack or other security failure.

2.3. Attackers

We assume that an attacker has the following capabilities:

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

An attacker can capture and replay signaling previously observed by it.

An 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 attack scenarios in which an attacker compromises intermediaries in the call path, or captures credentials that allow

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

This threat model also does not consider scenarios in which the operators of intermediaries or gateways are themselves adversaries who intentionally discard valid identity information (without a user requesting anonymity) or who send falsified identity using their own credentials. The design of the credential system will however limit the scope of the credentials issued to carriers or national authorities to those numbers that fall under their purview.

3. Attacks

The uses of impersonation described in this section are broadly divided into two categories: those where an attacker impersonates an arbitrary identity in order to disguise their own, and those where an attack will not succeed unless the attacker impersonates a specific identity. At a high level, impersonation encourages targets to answer attackers' calls and makes identifying attackers more difficult. This section shows how concrete attacks based on those different techniques might be launched.

3.1. Voicemail Hacking via Impersonation

A voicemail service allows users calling from their phones access to their voicemail boxes on the basis of the calling party number. If an attacker wants to access the voicemail of a particular target, the attacker may try to impersonate the calling party number using one of the scenarios described below.

This attack is closely related to attacks on similar automated systems, potentially including banks, airlines, calling-card services, conferencing providers, ISPs and other businesses that fully or partly grant access to resources on the basis of the calling party number. It would also be analogous to an attack where a human is encouraged to answer a phone, or to divulge information once a call is in progress, by seeing a familiar calling party number.

The envisioned countermeasures for this attack involve the voicemail system treating calls that supply an authenticated identity differently from other calls. In the absence of identity, for example, a voicemail service might enforce some other caller authentication policy (perhaps requiring a PIN for caller authentication). Authenticated identity alone provides a positive confirmation only when an identity is claimed legitimately; the

absence of authenticated identity here may not be evidence of malice, just of uncertainty.

If the voicemail service could learn ahead of time that it should expect authenticated identity from a particular number, that would enable the voicemail service to adopt stricter policies for handling a request without authenticated identity. Since users typically contact a voicemail service repeatedly, the service could for example

remember which users usually sign their requests and require further authentication mechanisms when signatures are absent. Alternatively, issuers of credentials or other authorities could provide a service that informs verifiers that they should expect identity signatures in calls from particular numbers.

3.2. Unsolicited Commercial Calling from Impersonated Numbers

The unsolicited commercial calling, or for short robocalling, attack is similar to the voicemail attack, except that the robocaller does not need to impersonate the particular number controlled by the target, merely some "plausible" number. A robocaller may impersonate a number that is not an assignable number (for example, in the United States, a number beginning with 0), or an unassigned number. A robocaller may change numbers every time a new call is placed, even selecting numbers randomly.

A closely related attack is sending unsolicited bulk commercial messages via text messaging services. These messages usually originate on the Internet, though they may ultimately reach endpoints over traditional telephone network protocols or the Internet. While most text messaging endpoints are mobile phones, increasingly broadband residential services support text messaging as well. The originators of these messages typically impersonate a calling party number, in some cases a "short code" specific to text messaging services.

The envisioned countermeasures to robocalling are similar to those in the voicemail example, but there are significant differences. One important potential countermeasure is simply to verify that the calling party number is in fact assignable and assigned. Unlike voicemail services, end users typically have never been contacted by the number used by a robocaller before. Thus they can't rely on past association to anticipate whether or not the calling party number should supply authenticated identity. If there were a service that could inform the terminating side of that it should expect an identity signature in calls or texts from that number, however, that would also help in the robocalling case.

When a human callee is to be alerted at call setup time, the time

frame for executing any 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 and network operations must be minimized for these countermeasures to be practical. For text messages, a delay for executing anti-impersonation countermeasures is much less likely to degrade perceptible service.

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 messages, or to force robocallers to use authenticated identity for numbers traceable to them, which would then allow for other forms of redress.

3.3. Telephony Denial-of-Service Attacks

In the case of telephony denial-of-service (or TDoS) attacks, the attack relies on impersonation in order to obscure the origin of an attack that is intended to tie up telephone resources. By placing constant telephone calls, an attacker renders a target number unreachable by legitimate callers. These attaacks might target a business, an individual or a public resource like emergency responders; the attack may intend to extort the target or have other motivations. Attack calls may be placed from a single endpoint, or from multiple endpoints under the control of the attacker, and the attacker may control endpoints in different administrative domains. Impersonation in this case allows the attack to evade policies that would block based on the originating number, and furthermore prevents the victim from learning the perpetrator of the attack, or even the originating service provider of the attacker.

As is the case with robocalling, the attacker typically does not have to impersonate a specific number in order to launch a denial-of-service attack. The number simply has to vary enough to prevent simple policies from blocking the attack calls. An attacker may however have a further intention to create the appearance that a particular party is to blame for an attack, and in that case, the attacker might want to impersonate a secondary target in the attack.

The envisioned countermeasures are twofold. First, as with robocalling, ensuring that calling party numbers are assignable or assigned will help mitigate unsophisticated attacks. Second, if authenticated identity is supplied for legitimate calls, then Internet endpoints or intermediaries can make effective policy decisions in the middle of an attack by deprioritizing unsigned calls when congestion conditions exist; signed calls, if accepted, have the

necessary accountability should it turn out they are malicious. This could extend to include, for example, an originating network observing a congestion condition for a destination number and perhaps dropping unsigned calls that are clearly part of a TDoS attack. As with robocalling, all of these countermeasures must execute in a timely manner to be effective.

There are certain flavors of TDoS attacks, including those against emergency responders, against which authenticated identity is unlikely to be a successful countermeasure. These entities are effectively obligated to attempt to respond to every call they receive, and the absence of an authenticated identity signature, or even the presence of an invalid signature, in many cases will not remove that obligation.

4. Attack Scenarios

The examples that follow rely on Internet protocols including SIP [1] and WebRTC.

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 SS7 call setup message (the CPN field of an IAM). When the call setup message reaches the terminating telephone switch, the terminal renders the attacker's chosen calling party number as the calling identity.

Impersonation, PSTN-PSTN

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

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 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.

4.1. Solution-Specific Attacks

Solution-specific attacks are outside the scope of this document. Some of the attacks that should be considered in the future include the following:

Attacks Against In-band

Token replay

Removal of in-band signaling features

Attacks Against Out-of-Band

Provisioning Garbage CPRs

Data Mining

Attacks Against Either Approach

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

Canonicalization attacks

5. Acknowledgments

David Frankel, Penn Pfautz, Stephen Kent, Brian Rosen, Alex Bobotek, Henning Schulzrinne, Hannes Tschofenig, Cullen Jennings and Eric Rescorla provided key input to the discussions leading to this document.

6. IANA Considerations

This memo includes no request to IANA.

7. Security Considerations

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

8. Informative References

[1] 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.

[2] Peterson, J., Schulzrinne, H., and H. Tschofenig, "Secure Telephone Identity Problem Statement", draft-ietf-stir-problem-statement-01 (work in progress), December 2013.

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

Author's Address

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

Email: jon.peterson@neustar.biz