Messaging, Malware and Mobile Anti-Abuse Working Group

M^{3}AAWG Telephony Honeypots: Benefits and Deployment Options

August 2014

Table of Contents

Foreword 1
Introduction 2
What Is a Honeypot? 2
Why Use Honeypots for Intelligence? 2
What Are the Benefits of Honeypots? 3
Implementation 4
Setting up a Telephony Honeypot 4
Types of Phone Numbers 5
Telephony Honeypot Deployment Scenarios 5
Scenario 1: Call Detail Record Only Honeypot 6
Scenario 2: Limited Interaction Honeypot 6
Scenario 3: Full Interaction Honeypot 7
Seeding of the Honeypot Phone Numbers 8
Scaling and Operational Considerations 9
Honeypot Data Usage 9
Research and Defense 9
Case Prioritization 9
Source Tracking 10
Feedback Loops 10
Partners 10
Authors 11

Foreword

In late 2013, the Messaging, Malware and Mobile Anti-Abuse Working Group (M^{3}AAWG) created the Voice and Telephony Abuse Special Interest Group (VTA-SIG) to address the growing abuse of public communications services that use a telephone number as an end address. This includes voice calls as well as text message services (SMS, MMS and RCS). The M^{3}AAWG VTA-SIG brings together experts from industry, government and academia to address telephony abuse including robocalling, voice phishing, caller-ID spoofing and telephony denial-of-service attacks.

Honeypots are a proven technology used for detecting and understanding online threats. A group of VTA-SIG participants started to investigate the potential benefits of telephony honeypots. To further explore this idea, several telephony honeypot instances have been created recently and early experiences with these honeypots indicate the value they offer in understanding and combating telephony abuse. Telephony honeypot development and use is now an active and growing effort in which several M^{3}AAWG members participate.

This document was written to facilitate and encourage honeypot development, as well as the use and sharing of information about and from honeypots. It includes an overview of the benefits of such honeypots and also provides details of the various options that exist for setting them up. We hope this document will help grow this effort and also provide an impetus for information sharing and collaboration to fight telephony abuse. The considerable experience M^{3}AAWG has in these areas that is shared in this document will undoubtedly be extremely valuable in maximizing the impact of others' efforts. We encourage you to construct and/or use honeypots and honeypot data, and invite you to contact M^{3}AAWG to collaborate with the VTA-SIG. (See http://www.m3aawg.org/vta-sig.)

Introduction

The frequency of unwanted calls (also called telephony spam) on our phones has increased at an alarming rate. The regulatory, law enforcement and service provider communities have received millions of complaints from citizens about unsolicited and fraudulent calls. The recent increase in attacks over the telephony channel can be attributed to the increased accessibility and anonymity provided by IP telephony, specifically Voice over Internet Protocol (VoIP). Bulk, automated and anonymous calls can be made from remote and concealed locations anonymously at no or low cost, creating a landscape for abuse similar to email spam. Criminals are exploiting the telephony channel to craft an increasing variety of attacks.

The most useful information about telephony spam widely available today comes from crowd-sourced complaint databases. Unfortunately these fall short of providing sufficiently complete, accurate, timely and sharable intelligence. To address these limitations, such information can be augmented by an alternative collection method known as "telephony honeypots." This document describes the application of telephony honeypots, detailing both the **types of data collected** and the use of such data in **analysis and mitigation of unwanted calls**.

What Is a Honeypot?

A telephony honeypot is a telephone service endpoint to which calls can be directed. It may appear to callers to be a normal telephone number (e.g., a typical 10-digit residential or business phone number), but is specifically designed and deployed to collect information on unwanted calls. It might automatically process calls or employ humans, is computer monitored and might be recorded.

Why Use Honeypots for Intelligence?

Currently there is limited visibility into overall telephony fraud and abuse due to the nature of how complaints from consumers are received and the evolving methods used by fraudsters to hide themselves. Some of the reasons for augmenting currently available information with intelligence collected from honeypots are as follows:

1. The Web-based platforms and protocols that are commonly used to collect telephony abuse reports are not easy to use for many victims who are not familiar with or do not have Internet access, including many of the most vulnerable people. A victim must know where to submit the complaint on the Web. Although sizable volumes of complaints are currently received by Web-based platforms[1], it is difficult to assess their *completeness*.

[1] http://www.ftc.gov/system/files/documents/reports/consumer-sentinel-network-data-book-january-december-2013/sentinel-cy2013.pdf

2. Due to the open nature of complaint collection, complaints are not limited to telephony fraud; people use complaint platforms to enter complaints about non-telephony abuse like email spam, junk mail or door-to-door solicitation. Typos or other inaccurate user input on Web forms also make complaints unreliable. This could result in spurious data that does not pertain to valid telephony abuse incidents which impacts the *accuracy* of such data.
3. There is a delay between when the fraudulent calls are made to people and when complaints are reported online by them. A complaint may be reported minutes, days or even weeks after an abusive call was made. Further, it is often impossible or impractical to discern the precise time or the number of unwanted calls received. Thus, it is difficult to ensure *timeliness* of user complaint data.
4. Honeypots can facilitate data *sharing* by minimizing privacy issues. For example, a cooperating Telecommunications Service Provider (TSP) could provide specific Call Detail Records (CDRs)[2] of incoming calls to selected honeypots on their network (i.e., a TSP-run honeypot), providing incremental intelligence to the investigators that could be used to create a backward link to the source of the call. However, these records are difficult to obtain today due to the privacy of the consumer information held by the carrier. In contrast, the owners of honeypot numbers have limited, to no, privacy concerns standing in the way of obtaining information about numbers they own because these numbers are not assigned to customers. With non-honeypot sources of information, privacy issues complicate data dissemination and fusion.
5. Miscreants and violators who are the source of telephony abuse make use of technologies that outpace Web complaint forms. For example, violators often use caller ID spoofing[3] to hide their phone number or misrepresent it to consumer victims, who then report the abuse using the spoofed number that appeared on their phone screen. It is estimated that 30% to 50% of the current volume of complaints involve an element of caller ID spoofing, thus reducing the value of actionable intelligence. Honeypots may identify technical similarities in spoofed calls, allowing correlation of seemingly random spoofed events that may not be possible with user complaint data.

What Are the Benefits of Honeypots?

There are several benefits of using a Honeypot Program, including a better understanding of the extent of telephony abuse, enforcement and investigative processes:

1. Honeypots are a "no consumer-required" mechanism for collecting intelligence on an ongoing and automated basis.
2. The research enabled by the Honeypot Program facilitates a data driven understanding of threats and facilitates research on countermeasures.
3. A program provides ongoing statistical data, trends and information about the evolving nature of abuse and generates awareness of unknown abuse case scenarios.
4. A program will discover potential violations that currently go unreported.
5. A honeypot will provide intelligence on which TSPs, regions and languages are being impacted and what form of violation is common.
6. Honeypots complement reporting mechanisms, permitting calibration of ratios of complaints to actual calls made (even with spoofing), facilitating an accurate estimate of actual calls made.

In summary, there is a mismatch today between systems that are used by violators and complaint collection technologies used by defenders. The fact remains that less tech-savvy consumers make up a significant

[2] http://en.wikipedia.org/wiki/Call_detail_record
[3] http://en.wikipedia.org/wiki/Caller_ID_spoofing

portion of the population and the violators of the law have significant abilities to hide their infrastructure and identities. A Honeypot Program can help address this mismatch and improve intelligence about telephony abuse.

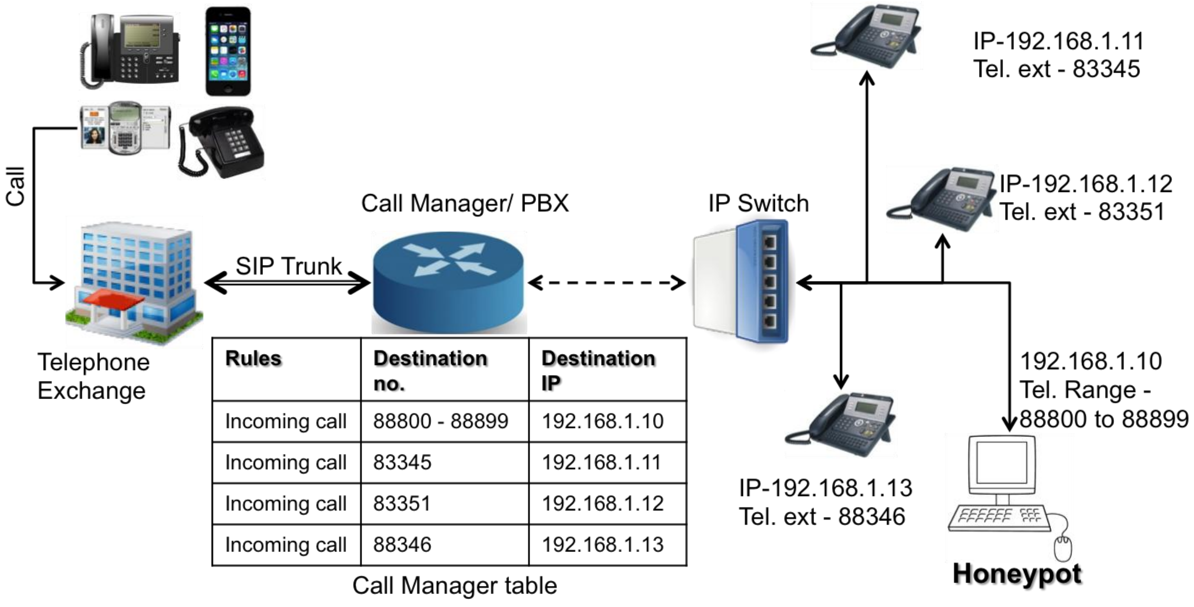
Implementation

This section provides details of what it takes to set up a telephony honeypot. This includes the infrastructure requirements and the telephone numbers that should be used. This document also explores some of the deployment scenarios of the honeypot. Some of the honeypot phone number seeding techniques to receive unsolicited calls from fraudsters are also highlighted.

Setting up a Telephony Honeypot

The build out of telephony honeypots is ideally accomplished by, or in collaboration with, a TSP that is able to provide phone numbers, call routing, call data records and/or backtrace information. Honeypots can be hosted inside or outside of TSP networks; for example, in a Private Branch Exchange (PBX). Low cost Internet call routing to a PBX may be implemented using Voice over Internet Protocol[4] communications. A host PBX may be constructed using open source software-based solutions, such as Asterisk[5], and server applications to accept, route, interact and log the incoming connections. Other implementation options for hosting the honeypot infrastructure are discussed further in this document.

A reference model is shown in Figure 1, illustrating the connectivity and logical architecture in which a PBX provides end routing to a Honeypot.



Rules	Destination no.	Destination IP
Incoming call	88800 - 88899	192.168.1.10
Incoming call	83345	192.168.1.11
Incoming call	83351	192.168.1.12
Incoming call	88346	192.168.1.13

Figure 1: Telephony Honeypot Setup Using 1

The TSP routes calls to an Internet Protocol Public Branch Exchange (IP-PBX[6]). As calls are received, various actions would be performed as outlined in the Telephony Honeypot Deployment Scenarios section of this paper.

[4] http://en.wikipedia.org/wiki/Voice_over_IP
[5] http://www.asterisk.org/
[6] http://www.asterisk.org/get-started/applications/pbx

Types of Phone Numbers

There are several types of phone numbers that can be requested from TSPs:

1. Numbers that have never been assigned
    a. New area codes or NXXs[7] within an area code
    b. Complete new NXXs may also be provided

2. Reused/recycled phone numbers that are not assigned and are of varying issue date
    a. Older numbers are more valuable to miscreants as they demonstrate a long term billing address and potentially an elderly person
    b. Other reused phone numbers with specific attributes that TSPs can provide

3. Dirty numbers – Numbers which have been given up by customers of a TSP or a service provider because of a significantly high number of unsolicited incoming calls

4. Commercial phone numbers
    a. Retirement homes are of particular interest

5. Governmental phone numbers
    a. Discontinued governmental phone numbers

6. Sequencing of numbers will also be of value to discover sequential dialers and the range of impact of specific ADAD (Automatic Dialing-Announcing Device) campaigns.

The program should account for constant churn in phone numbers provided by TSPs on an ongoing basis.

Telephony Honeypot Deployment Scenarios

This paper classifies honeypot deployment scenarios into three broad categories based on the interaction level with the attackers: a) CDR-Only Honeypot, b) Limited Interaction Honeypot and c) Full Interaction Honeypot. Apart from the first category, each scenario can be deployed with or without the voice recording feature. The attacker makes calls to the "honeypot numbers" included in the overall pool of numbers in their dial-out campaigns. Data collected in the honeypot – e.g. CDRs, voice recordings, etc. – would subsequently be stored to build intelligence reports of calls coming into honeypot numbers within a TSP network. The retention policy for storing these transactions should be consistent with other similar data types collected.

[7] The three-digit exchange code, or central office code, that appears after the area code in a phone number.

Scenario 1: Call Detail Record Only Honeypot

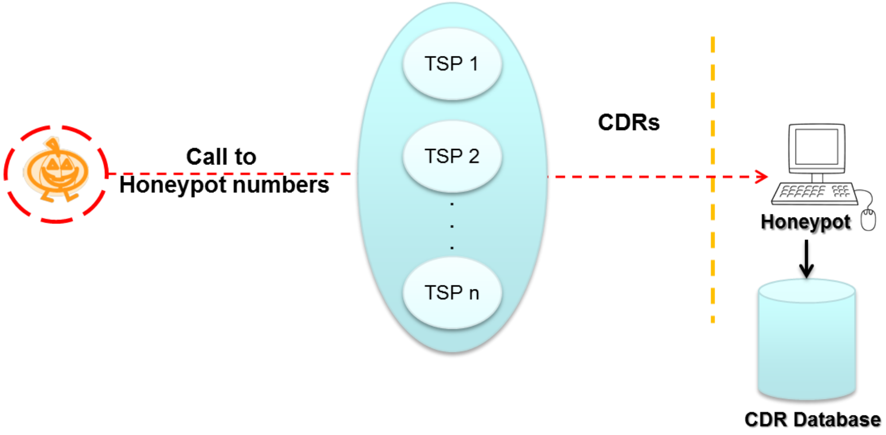


Figure 2: Call Detail Record Only Honeypot

In this scenario, TSPs do not provide call routing functionality and calls to the honeypot phone numbers are terminated within their own network (see Figure 2). Various call completion behaviors, such as "not in service" recordings, no answer, busy, etc., may be considered. After the call is made, the TSP would provide the Call Detail Records to the operator of the honeypot using a secure protocol such as a SSL or VPN connection. A CDR can be accompanied with additional information about how the call was handled (e.g., busy, seize the hangup, no answer or "not a working number" response).

Scenario 2: Limited Interaction Honeypot

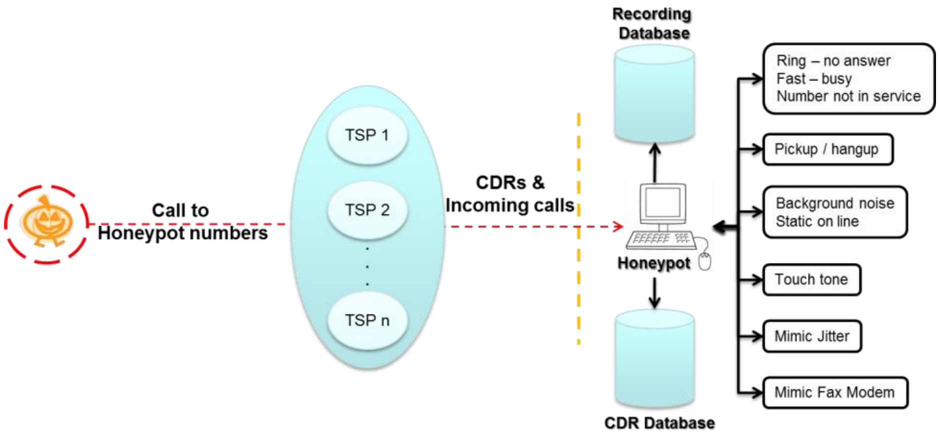


Figure 3: Limited Functionality Incoming Call Honeypot

In a limited interaction scenario, TSPs provide call routing functionality and calls to the honeypot phone numbers are terminated at the honeypot-provided infrastructure (see Figure 3). Call completion behaviors include non-termination (such as "not in service" recordings, no answer, busy, etc.) that might be considered.

Some of the options are as follows:

1. Do not answer the call – A portion of calls are routed to a ring no-answer, fast-busy, or number-not-in-service type message.
2. Pick-up/hang-up – A portion of calls are picked up and immediately hung-up using an application on the honeypot server.
3. Noise-only connection – A portion of calls are picked up via a server application with only white noise or similar background noise such as "static on the line" played for 60 seconds or so in various random volumes, patterns and/or duration.
4. DTMF/touch tones – A portion of calls are picked up via a server application on the honeypot to mimic the consumer dialing DTMF (Dual Tone Multi-Frequency) tones; e.g., "Dial 1 to continue."
5. Injection of Jitter – Jitter and latency are added to a portion of the calls to mimic poor connectivity.
6. Fax-modem – A fax-modem tone is played back to mimic a fax-modem connection for a portion of the calls.

Scenario 3: Full Interaction Honeypot

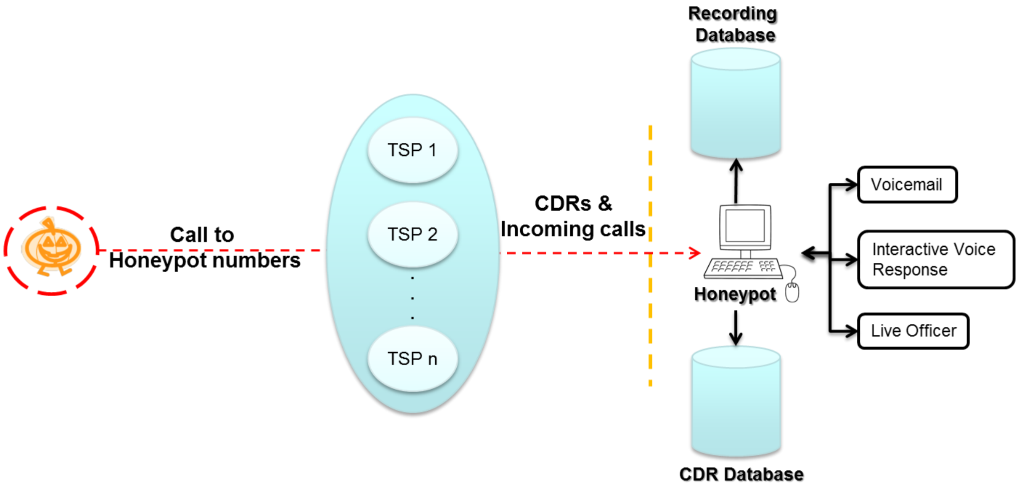


Figure 4: Voicemail Interaction Honeypot with Recording

There are three levels of interactivity of a honeypot (see Figure 4):

1. Voicemail - A portion of calls are **answered and recorded** using voicemail announcements that indicate the calls are being recorded. These recordings are subsequently automatically processed using a variety of technologies to determine the origin of the call[8], type of call (ADAD or live voice) and other heuristic data relevant to the honeypot infrastructure. In all recording scenarios, an announcement indicating that the call is being recorded should be played before the recording commences. Announcement voices can be randomly generated with varying pitch, volume and voice annotations or other factors by making use of computer-generated voices. For example, a honeypot of 10,000 voice mailboxes can offer a range of 9,000 voice patterns, providing potential violators the sense that each mailbox is unique. Other standard or generic voicemail greetings generated by the TSP

[8] http://www.pindropsecurity.com/

that indicate the call is being recorded can be used. In these scenarios, identical recording durations, including warnings of the message limits as well as other interactive voicemail features, would be played through a TSP voicemail system.

2. Interactive voice response (IVR) – A portion of calls are answered and recorded via an IVR system with a clear indication that the call is being recorded.
3. Live person – Where the caller ID is under investigation, the call could be routed to an assigned person, such as a compliance officer or law enforcement officer, to allow for interaction with the suspected miscreant. Calls will be recorded in this case and the other party should be informed of the recording at the start of the call. Routing of calls to an officer could occur at random as well.

Seeding of the Honeypot Phone Numbers

Seeding of a honeypot phone number refers to how a phone number is advertised to make it attractive to fraudsters. Unlike email addresses, the total potential phone number space is limited and it is possible that fraudsters will choose targeted phone numbers randomly or sequentially. On the other hand, to increase their success rate, they may qualify targeted phone numbers with information scraped from various sources. To understand this process, honeypot phone numbers must be seeded systematically.

To understand the affect of seeding phone numbers and whether specific attackers are picking their targets randomly or are using some kind of qualification method, some phone numbers should never be seeded anywhere. Then a portion of the remaining honeypot phone numbers should be seeded at various places on the Web with the assumption this will attract scammers but not legitimate users. Possible places to seed phone numbers include:

1. Online Social Networks: Online social networking sites like Facebook, Twitter and personal Web blogs or websites could potentially be targets for scammers to scrape and obtain phone numbers. Phone numbers can also be posted as comments on existing popular sites and blogs. For example, phone numbers can be tweeted on Twitter feeds or added to the comments of most popular or trending Facebook posts.
2. Questionable websites: A few website categories are more likely to be scraped by scammers, such as drugs, gambling and adult dating, and porn or sex related sites. Most of these sites are known to be spam or malware intensive and thus come under the classification "questionable."
3. Call-To: There are a large number of websites publishing suspicious toll free numbers; e.g., http://800notes.com/Phone.aspx/1-800-931-1026. Making calls to those toll-free services from honeypot lines is another approach to seeding phone numbers. Moreover, calls can be made to the fraud and reported numbers posted on 800notes[9] and on other regulatory and law enforcement complaint databases. This will work only when numbers reported at these sites are not spoofed.
4. Miscellaneous: Phone numbers can also be posted at some other websites which are likely to be the target of scammers, such as websites related to advertising, travel, insurance, finance, etc.

[9] http://800notes.com - a free reverse phone number lookup and database of end-users' submissions on fraudulent and annoying calls.

Scaling and Operational Considerations

Unfortunately telephone numbers are a limited rather than disposable resource; it is not practical to routinely abandon honeypot telephone numbers that may have been rendered unfit for normal use by seeding or other activities. Therefore, lifecycle issues need to be considered in scaling and seeding plans.

Honeypots will loose value if attackers are able to identify them. Attackers may then avoid calling them, provide false or misleading information to disrupt defenses, or otherwise interfere with the operation of honeypots. Therefore, information that may be used to identify them, including but not limited to their existence, IP addresses and telephone numbers, should be protected accordingly.

Honeypot phone numbers can be incrementally or gradually provisioned or rotated with a pool of inactive numbers by an operator to allow for capacity management and monitoring of systems supporting the implementation scenario. This will also ensure freshness and help prevent attackers from evading known honeypots.

As honeypots become known as an integral component of telephone system defenses, they may become targets of focused Telephony Denial of Service Attacks[10] (TDoS). These attacks may disrupt not only the honeypots, but also other critical telephone and Internet services. It is therefore prudent to establish plans and procedures, such as granting a TSP permission to shutdown the honeypot phone numbers being targeted as needed.

Honeypot Data Usage

Data collected from honeypots would ideally be used not only by the honeypot operator but also by other organizations. Data sharing between honeypot operators, service providers and enforcement organizations allows for a more comprehensive view of abuse, and the fusion of data from multiple sources facilitates more effective mitigation. Several of these uses are discussed below.

Research and Defense

The current methodology of collecting information on unsolicited calls does not provide accurate and real time information. Data collected from a honeypot can help telephone security researchers understand the nature of the telephony threat and its scale. This data will also help researchers develop techniques to analyze the threat early and takedown multiple sources. For example, CDRs from the honeypot can be used to detect several calling patterns from telemarketers or debt collectors who employ aggressive and possibly illegal tactics.

Case Prioritization

Law enforcement and regulation bodies lack the resources to investigate every call that is potentially fraudulent. The data collected from the honeypot can be used by such agencies in combination with consumer Web-form submitted complaints and information gathered during inspections to help prioritize cases. This dataset can add timely and more accurate information then Web-form submitted complaints alone.

As the numbers and seeding of the honeypot mature, the detected calling patterns can be used to help roughly predict how many calls a particular number made, based on the broadness of distribution.

[10] http://en.wikipedia.org/wiki/Denial-of-service_attack#Telephony_denial_of_service

Source Tracking

There are several attributes of a voice call that can be used within the context of a honeypot to track down the source of a potential violation.

1. Call Detail Records from systems might highlight which carriers are most impacted by certain campaigns.
2. The IP address of the SIP server connecting to the honeypot, as well as those that may be present in SIP header information, can be used to help determine the source.

Recorded sounds from the call can be used to help determine precise location of the potential violator, type of infrastructure being used and several other data sets unavailable with today's investigation techniques. For example, phone printing[11] technologies could be used for this.

Feedback Loops

An originating service provider may be unaware of or have insufficient basis for enforcement activities against an abuser. For example, robocalls generated from Service Provider 1's network may target phones in Service Provider 4's network, routed through the networks of service providers 2 and 3. Service providers 1, 2 and 3 may be unaware of the problem.

Service Provider 4 may have honeypots or other abuse sensors. By providing abuse indications to Service Provider 1, Service Provider 4's honeypot data may empower enforcement activities at the source, within Service Provider 1's network. Provided adequate trust between service providers and the ability to share honeypot data, this sort of collaboration can benefit the entire ecosystem, and with broad sharing, can even be effective in mitigating spoofed calls.

For additional information on data sharing, the IETF RFC 6449, Complaint Feedback Loop Operational Recommendations, November 2011[12], is an excellent reference. Although written specifically for email, nearly all of its principles are applicable to phone spam.

Partners

Honeypots can be deployed with the help of the following types of organizations that may host the infrastructure in-part or completely to support the program:

1. Nonprofit Organizations such as M^{3}AAWG, NCFTA U.S. or NCFTA Canada
2. Universities
   a) Georgia Institute of Technology and the New York University Abu Dhabi currently have set up a telephony honeypot with a limited set of phone numbers. Several other educational institutions around the globe are exploring the deployment of such honeypots.
   b) Concordia University is participating in the Canadian telephone honeypot program.
3. Telephone Service Providers
4. Telephone Consulting Service Companies via the Request for Proposal process
5. Telephony Security Vendor Companies
6. Various Regulatory Agencies, such as the CRTC[13], the U.S. Federal Communications Commission (U.S. FCC) and the U.S. Federal Trade Commission (U.S. FTC)

Inquiries for additional information can be sent to vtasig-chair@m3aawg.org and telephony_honeypot@gtisc.gatech.edu.

[11] http://www.pindropsecurity.com/about-phone-printing/
[12] http://tools.ietf.org/html/rfc6449
[13] (Canadian Radio-television and Telecommunications Commission)

Authors

Payas Gupta, New York University Abu Dhabi
Mustaque Ahamad, Georgia Institute of Technology and New York University Abu Dhabi
Jonathan Curtis, Canadian Radio-television and Telecommunications Commission[14]
Vijay Balasubramaniyan, Pindrop Security
Alex Bobotek, Messaging, Malware, Mobile Anti-Abuse Working Group (M^{3}AAWG)

[14] *The views expressed herein are those of the author and do not necessarily reflect the position of the CRTC or the Government of Canada.*