METHODS FOR COMBATING MESSAGING FRAUD

SESSION APP-1103
Agenda

• Introduction
  Problem Statement
  Goals (and Non-Goals) of Message Authentication
  Terminology

• Solution Classes
  Path-Based (SPF, Sender ID)
  Signature-Based (DKIM)
  Miscellaneous (CSV, BATV)

• Deployment
  Software Availability
  Publishing and Verification of Records

• Q and A
What Is Messaging Fraud?

- Fraud relating to the source or description of a message
  
  “Message” may be e-mail, IM, or telephony

- Content may be fraudulent too
  
  Difficult to automatically interpret content to detect fraud
  
  Some specialized detection tools available
Message Authentication

Message Authentication is the process of determining whether a message is actually from its claimed source.
Why Message Authentication?

- Makes it harder to hide—improves accountability
- Ability to claim message came from another is integral to some fraud schemes
  - Many types of phishing
  - Confidence schemes
- Want to improve users’ trust of the Internet
- Required to support reputation and accreditation of senders
What Message Authentication *Won’t* Do

- **Solve the spam problem**
  
  Spammers will send messages from throwaway addresses
  
  Accountability is limited by that of domain registration

- **Solve the phishing problem**
  
  Human-engineered and look-alike domain names still exist
  
  alerts@bigbank-security.com, fraud@example.com
Emerging Attack Vectors

- Domain names are being internationalized
  
  Unicode characters can be used to represent domains using non-ASCII alphabets
  
  These characters are sometimes hard to distinguish from ASCII counterparts

- Internationalized domain names can look like other familiar domains

  security@bigbank.com

  🇷🇺 Cyrillic small letter a (Unicode 0430)

- Ambiguity is often font-dependent
Some Terminology

- **MTA (Mail Transfer Agent)**—a “mail server”
- **MUA (Mail User Agent)**—what a user uses to send/receive mail
- **Message Envelope**—addressing information sent with a message
- **Transparent forwarder**—an MTA that resends a message without modification except to the envelope
- **Phishing**—use of e-mail to lead consumers to counterfeit websites *(source: Anti-Phishing Working Group, antiphishing.org)*
What Is a Message’s Source?

MAIL FROM
From:
Screen Name
HELO/EHLO
Resent-From:
P-Asserted-Identity:
Resent-Sender:
Sender:
Message Source

• A number of different “sources” may exist, with different semantics

• The semantics are inconsistently used, especially for email
  
  Mailing lists add/modify headers differently
  Different client software adds/uses different headers
  Some header semantics rarely used, but “standard”
Source Address Characteristics
From/Sender

• From:
  Most frequently displayed to recipient
  Outlook typically displays only the “friendly” address, a problem for “John Chambers” <biff@hacker.com>
  Rarely used, but From can contain multiple addresses, e.g.,
    From: <castor@twins.org>, <pollux@twins.org>
  Sender: indicates origin in this case

• Sender:
  Indicates who injected the message
  Mailing lists are considered an injection, so many rewrite or add Sender
Source Address Characteristics (Cont.)

Resent Fields

- Used to indicate messages that have been reinjected into the mail system
- Resent-From:
  Address(es) which reintroduced the message
- Resent-Sender:
  Specific address that reintroduced the message
- Since messages can be resent more than once, multiple blocks of Resent headers may exist
Source Address Characteristics

Envelope From

- Also referred to as “MAIL FROM”, “2821 From”
- Address to which bounce messages should be sent
- Null if the message is already a bounce
  Don’t send bounces in response to bounces
- May be rewritten by mailing lists
  Particularly if list owner should get the bounce messages
- Not usually rewritten by transparent forwarders
  Unless ultimate recipient wants anonymity
- May be an unrelated address used to track bounces
  Particularly used by some commercial bulk mailers
Source Address Characteristics
HELO/EHLO Domain

- Characteristic of the sending MTA, not the message itself
- MTA identity is often significant
  Good mail tends to come from good MTAs
  And vice versa (zombies?)
- Allows name of the MTA to be determined and verified (without reverse DNS)
- Frequently mis-implemented or misconfigured in MTAs, currently with little effect
  Should say “HELO <my name>”
  But often say “HELO <your name>”
Agenda

• Introduction
  Problem Statement
  Goals (and Non-Goals) of Message Authentication
  Terminology

• Solution Classes
  Path-Based (SPF, Sender ID)
  Signature-Based (DKIM)
  Miscellaneous (CSV, BATV)

• Deployment
  Software Availability
  Publishing and Verification of Records

• Q and A
PATH-BASED TECHNIQUES
Path-Based Techniques: Introduction

• Philosophy:
  “All messages should come from an MTA authorized by the sending domain”

• Predominant technologies
  Sender Policy Framework (SPF)
  Sender ID

• SPF and Sender ID differ primarily on the message identity they use
  SPF uses MAIL FROM address
  Sender ID uses selected header known as “Purported Responsible Address” (PRA)
Path-Based Techniques: General

Would I expect a message from this sender to come via a.b.c.d?
How Path-Based Methods Work

- After (or if possible, during) receipt of a message, determine its origin address
  
  Choice of origin address depends on method being used

- Look for a DNS TXT resource record from the originating domain

- If record exists, it gives information on outgoing mail servers used by that domain
## SPF Record Format

<table>
<thead>
<tr>
<th>v=spf1</th>
<th>Identifies the record as an SPF (version 1) record</th>
</tr>
</thead>
<tbody>
<tr>
<td>a[:host.example.com]</td>
<td>Address the domain resolves to [or host.example.com resolves to] is valid</td>
</tr>
<tr>
<td>mx[:example.com]</td>
<td>Addresses corresponding to this domain’s [or example.com’s] Mail Exchanger records are valid</td>
</tr>
<tr>
<td>ptr[:example.com]</td>
<td>Addresses which reverse-resolve to an address in this domain [or example.com’s domain] are valid</td>
</tr>
<tr>
<td>ip4:a.b.c.d/m</td>
<td>The address or subnet defined by a.b.c.d with netmask m are valid</td>
</tr>
<tr>
<td>include:example.com</td>
<td>Addresses permitted by example.com are valid</td>
</tr>
</tbody>
</table>
# SPF Failure Types

<table>
<thead>
<tr>
<th>?all</th>
<th>An SPF failure should be considered “neutral” (no information)</th>
</tr>
</thead>
<tbody>
<tr>
<td>~all</td>
<td>Softfail: Messages should not be rejected on failure, but may be subjected to added scrutiny</td>
</tr>
<tr>
<td>-all</td>
<td>Hardfail: Messages may be rejected or subjected to added scrutiny</td>
</tr>
</tbody>
</table>
Sample SPF Record

Cisco’s SPF Record:

- `cisco.com.  IN  TXT  “v=spf1 ptr a:mustang1.netsolve.com ~all”`

<table>
<thead>
<tr>
<th>Field</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>v=spf1</td>
<td>This is an SPF version 1 record</td>
</tr>
<tr>
<td>ptr</td>
<td>Any address which resolves to *.cisco.com is acceptable</td>
</tr>
<tr>
<td>a:mustang1.netsolve.com</td>
<td>Mail may also come from mustang1.netsolve.com</td>
</tr>
<tr>
<td>~all</td>
<td>Mail from other addresses should be treated with caution</td>
</tr>
</tbody>
</table>
Purported Responsible Address (PRA)

- Used by Microsoft’s Sender ID proposal to determine origin address
- Headers searched (approximate priority order):
  - Resent-Sender
  - Resent-From
  - Sender
  - From
- Details available at:
- Microsoft claims patent rights to the algorithm, but is licensing it under liberal terms
The Transparent Forwarding Problem

Message Sent to user@alum.college.edu
Gets Forwarded to “real” Recipient Address

Sender

Originating MTA

alum.college.edu

Recipient MTA

Recipient

Would I expect a message from this sender to come via e.f.g.h?

IP a.b.c.d

IP e.f.g.h

DNS

Recipient Sees Message as Coming from an Unexpected IP Address
DomainKeys Identified Mail (DKIM)

- DKIM is a hybrid of two prior message signature proposals
  - Identified Internet Mail (Cisco)
  - DomainKeys™ (Yahoo!)
- Header-based signature intended to protect sender from spoofing, cut-and-paste attacks, etc.
- Minimizes changes to transport infrastructure between signer and verifier

DomainKeys is a trademark of Yahoo! Inc.
DomainKeys Identified Mail Explained

1. Sender’s MTA Signs Message
2. Receiver’s MTA Verifies Message
3. DNS Provides Public Key and Its Authorization
4. Optional: Consult 3rd Party Reputation Service

Sending Domain

Receiving Domain
DKIM Characteristics

• Signature appears as an additional message header
  Generally ignored by non-signature-aware elements

• Signing and verification typically take place at MTAs, but may occur at MUA
  May occur at any point within the trust domain of originator and recipient

• PGP signature over selected headers and body
  Canonicalization may be used to allow “safe” modifications like spacing changes
Authentication/Authorization Model

Messages Must Pass Two Tests Before They Are Authenticated

AUTHENTICATE THE MESSAGE

Receiving Domain Authenticates the Message—i.e. Verifies that the Message Was Not Altered in any Consequential Manner Prior to Reaching the Receiving Domain

AUTHORIZE THE SENDER

Receiving Domain Asks Sending Domain to Confirm that Whoever Signed the Message Was Authorized to Do So (Without Having to Identify the Sender)
Example of DKIM Signed Message

Subject: Sample message
From: John Doe <jdoe@example.com>
To: Mary Smith <msmith@example.net>
Content-Type: text/plain
Message-Id: <1098727240.13184.0.camel@lucid.example.com>
Mime-Version: 1.0
X-Mailer: Ximian Evolution 1.4.6 (1.4.6-2)
Date: Wed, 25 May 2005 11:00:40 -0700
Content-Transfer-Encoding: 7bit
DKIM-Signature: a=rsa-sha1; d=example.com; s=may2005;
   i=jdoe@example.com; c=nowsp; q=dns; t:1098727241; x:10988893641;
   h=Subject:From:Date;
   b=QQgUTUMvDA1BPxxIpSrAiAUXB5rtOt4tJT1BcN3zB01pUARhybDLGF7KLU7ens
   WielZcm7+h5lfOhYvuy3DUTQ==;

Did you receive today’s sales orders yet?

- John
## What’s in a DKIM Signature?

<table>
<thead>
<tr>
<th>Tag</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
<td>Version (default = DKIM1.0)</td>
</tr>
<tr>
<td>a</td>
<td>Algorithm, e.g., rsa-sha1</td>
</tr>
<tr>
<td>b</td>
<td>Signature data</td>
</tr>
<tr>
<td>c</td>
<td>Body canonicalization, e.g., nowsp (default = simple)</td>
</tr>
<tr>
<td>d</td>
<td>Domain of signer</td>
</tr>
<tr>
<td>h</td>
<td>Signed headers</td>
</tr>
<tr>
<td>i</td>
<td>Identity associated with signature</td>
</tr>
<tr>
<td>q</td>
<td>Key query method(s) (default = dns)</td>
</tr>
<tr>
<td>s</td>
<td>Selector specifying key to use</td>
</tr>
<tr>
<td>t</td>
<td>Signature timestamp</td>
</tr>
<tr>
<td>x</td>
<td>Signature expiration time</td>
</tr>
<tr>
<td>z</td>
<td>Copied headers</td>
</tr>
</tbody>
</table>
What’s in a DKIM Key Record?

<table>
<thead>
<tr>
<th>Tag</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
<td>Version (default = DKIM1.0)</td>
</tr>
<tr>
<td>g</td>
<td>Granularity of key (user or all users)</td>
</tr>
<tr>
<td>k</td>
<td>Key type (default = rsa)</td>
</tr>
<tr>
<td>n</td>
<td>Human-readable notes</td>
</tr>
<tr>
<td>p</td>
<td>Public key data</td>
</tr>
<tr>
<td>s</td>
<td>Service type (default = any)</td>
</tr>
<tr>
<td>t</td>
<td>Flags, e.g., testing</td>
</tr>
</tbody>
</table>

- Records are stored in DNS TXT RRs at `selector._domainkey.example.com`
- Alternative RR types being discussed
Third-Party Signatures

- Sometimes a signature on behalf of other than the originator is useful/necessary
- Mailing lists need to sign when they modify messages
  
  May also want to sign to indicate that message came through the mailing list
- Some services like Evite want to send messages on behalf of users, but will sign on its own behalf
- **Risk**: Messages may be signed by attacker “on behalf of” someone else without their authorization
- **Mitigation**: Attempt to display signer’s identity to recipient if different from originator
Message Signing Policy

• How should unsigned mail from the domain be handled?
  
  If all messages are signed, unsigned ones are probably bogus
  
  Otherwise they may be acceptable

• May also want to limit re-signing by third parties
  
  Some senders are more interested in security than, for example, ability to traverse mailing lists
Deploying Message Signing

• Deploy a signature-capable MTA
  Major MTA appliance vendors are adding signature support
  “Milter” API software available for sendmail
  DomainKeys toolkit for other MTAs (e.g., qmail)

• Generate and publish message signing keys
  Published in DNS records in a separate subdomain
  May delegate key subdomain to mail administrators
  Optional: publish a message signing policy

• Tell users how to handle message verification results
CERTIFIED SERVER VALIDATION (CSV)
Introduction to CSV

• Philosophy: The reliability of the mail server correlates well with the desirability of the messages it sends

• Mail server identity is expressed in HELO/EHLO string

• CSV includes:
  
  DNS authentication of HELO string (does HELO string translate to the address in use)
  
  DNS-based authorization mechanism (CSA)
  
  Accreditation mechanism (DNA)
CSV Usage Example

Is this the IP of mail.example.org?
Is this an authorized outgoing mailer?
Are they reliable?

Sender → Originating MTA
“HELO mail.example.org”
Recipient MTA → Recipient

Sender → DNS
Recipient → Accreditor’s DNS
CSV Issues

• Many commercial MTAs mis-populate the HELO string because it doesn’t currently matter
  Receivers can’t interpret a bogus HELO string as fraud

• Relatively large dependence on accreditation or reputation systems
COMPARING MESSAGE AUTHENTICATION APPROACHES
Common Characteristics

- All depend on DNS integrity
  - Theoretically insecure, but good in practice (so far)
  - DNSSEC is coming...someday
- None are 100% reliable
  - No single approach always works
  - Goal of rejecting messages is difficult to achieve
  - Much easier to make a positive than a negative assertion about a message
## Comparison Matrix

<table>
<thead>
<tr>
<th></th>
<th>SPF</th>
<th>Sender ID</th>
<th>DKIM</th>
<th>CSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifies message before acceptance</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Survives transparent forwarding</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Minimal deployment requirements for sender</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>User-level granularity</td>
<td></td>
<td></td>
<td>✔</td>
<td></td>
</tr>
<tr>
<td>Mitigates message replays</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Deployable within recipient network</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Some effectiveness before reputation deployed</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
Comparison Comments

• There is no clear winner or loser
• Authentication methods are complementary
  Strengths of some are weaknesses of others
Agenda

• Introduction
  Problem Statement
  Goals (and Non-Goals) of Message Authentication
  Terminology

• Solution Classes
  Path-Based (SPF, Sender ID)
  Signature-Based (DKIM)
  Miscellaneous (CSV, BATV)

• Deployment
  Software Availability
  Publishing and Verification of Records

• Q and A
DEPLOYING MESSAGE AUTHENTICATION
What’s Required to Deploy?

- All approaches require sender to publish some data in DNS
- Signature-based approaches (DKIM) additionally require signer to compute and attach signature
- All require software at the verifier to evaluate the message
- Typically a verification header is added to indicate the results downstream

All approaches but DKIM need to be evaluated at the edge of the recipient domain
Software Availability

• Typically implemented by a plug-in to popular MTAs (e.g., sendmail’s milter API)

• Open-source code available for several popular schemes
  
  Sender ID milter available on Sourceforge
  
  DKIM code under development, soon to be released
  
  Exim patch for CSV
Resources: Technologies

• Sender Policy Framework
  http://spf.pobox.com/

• Sender ID
  http://www.microsoft.com/mscorp/safety/technologies/senderid/default.mspx

• DomainKeys
  http://antispam.yahoo.com/domainkeys

• Identified Internet Mail
  http://www.identifiedmail.com/

• Certified Server Validation
  http://www.mipassoc.org/csv/
Resources: Mailing Lists

- IETF message authentication signature standards
- SPF discussion
  http://archives.listbox.com/spf-discuss@v2.listbox.com/
- SPAM-L mailing list
  http://peach.ease.lsoft.com/archives/spam-l.html
- IETF-Clear mailing list (CSV, etc.)
  http://mipassoc.org/mailman/listinfo/ietf-clear
Resources: Organizations

- Messaging Anti-Abuse Working Group (MAAWG)
  
  http://www.maawg.org/

- APWG
  
  http://www.antiphishing.org/

- ASRG
  
  http://asrg.sp.am/

- IETF
  
  http://www.ietf.org
Q and A
Complete Your Online Session Evaluation!

- Win fabulous prizes! Give us your feedback!
- Receive 10 Passport Points for each session evaluation you fill out
- Go to the Internet stations located throughout the Convention Center
- Winners will be posted on the Internet stations and digital plasma screens
- Drawings will be held in the World of Solutions
  - Monday, June 20 at 8:45 p.m.
  - Tuesday, June 21 at 8:15 p.m.
  - Wednesday, June 22 at 8:15 p.m.
  - Thursday, June 23 at 1:30 p.m.