

# 追踪垃圾邮件来源

( v1.0 )

Written By: Refdom

Date: 2004-5-10

<http://www.xfocus.org>

声明: 安全焦点 (xfocus security team) 是非商业, 全方位的网络安全组织, 本文档为安全焦点发布的技术文件, 供自由技术传播, 拒绝商业使用。文档的所有权归属安全焦点, 任何使用文档中所介绍技术者对其后果自行负责, 安全焦点以及本文档作者不对其承担任何责任。

## 目录

1	简介 .....	3
2	邮件头分析 .....	3
3	邮件传输过程 .....	5
4	邮件头分析实例 .....	7
5	邮件伪造 .....	10
6	垃圾邮件分析 .....	11
	实例分析一 .....	11
	垃圾邮件实例二 .....	12
	垃圾邮件实例三 .....	15
7	总结 .....	16
8	参考 .....	17

## 1 简介

电子邮件是最常用的网络应用之一，已经成为网络交流沟通的重要途径。但是，垃圾邮件（spam）烦恼着大多数人，近来的调查显示，93%的被调查者都对他们接收到的大量垃圾邮件非常不满。现在世界上超过50%的电子邮件都是垃圾邮件，但是只有少数组织承担责任。

垃圾邮件也造成了很有影响的安全问题，比如 Phish 形式的欺诈。日益增加的垃圾邮件每年都会造成 20 亿美元以上的损失。

针对垃圾邮件造成的问题越来越引起了人们的重视，在对付垃圾邮件方面，主要有两种形式：防御与追踪。防御主要是强调对垃圾邮件的过滤或者是阻止垃圾邮件的产生，而追踪则强调主动地追查垃圾邮件来源，并对其进行警告或者采取其他措施。

实际上，不仅仅是垃圾邮件，现在，很多的网络安全事件都会同邮件直接关联，比如病毒传播、社会工程学、木马甚至一些反动的对国家安全造成危害的，都会通过邮件途径进行散播。对于这类事件，追踪来源将显得尤其重要了。

本文将主要介绍对邮件的追踪方面的技术，主要以垃圾邮件为例。通过针对邮件头进行分析，并查询到最接近源头的地址，这些技术并不一定能查询到最原始的源头，因为 spammer 通常会修改邮件头以便能够隐藏自己。

## 2 邮件头分析

追踪邮件将很大程度依靠对邮件头的分析，[RFC2076](#) 列出了多数通用的消息头，也可以参考 [RFC2822](#)。

这里是介绍一小部分，这些部分可能能帮助我们进行分析：

**From:**

邮件从哪里发送的。很容易被伪造，在分析中，非常不可信任。

**From**

不同于 From:域, 这行并不通常是邮件头的一部分, 但是邮件转发程序经常插入这一行, 表明邮件什么时候被接收的。这一行总是邮件头的第一行, 也可以被伪造, 但并不一定。

**Reply-To:**

回复时发送的地址。很容易被伪造, **但常常提供线索**, 比如有些垃圾邮件经常用该域指向一个合法的邮件地址, 以便 spammer 能够接收到回复的邮件。

**Return-Path:**

与 Reply-To:相同

**Sender:**

消息发送者。这通常都是伪造的

**In-Reply-To:**

在回复的时候可能存在, 通常指向原邮件的 Message-ID:

**Message-ID:**

邮件系统在创建邮件时的唯一标记 (参考 [RFC822](#) 、 [RFC1036](#))。也经常被伪造, 但如果是正常的, 那么 Message-ID:也通常能确定发送者所登录的系统, 而不仅仅是邮件被创建的系统。Message-ID 的结构同邮件服务器程序有直接关系, 不同的邮件服务器产生的 ID 也不一样, 有时相同邮件服务器的不同处理也会产生不一样的 ID, 但是, 多数邮件服务器都包含下面的一些内容:

- 日期
- 时间
- 一个唯一标识

- DNS

有的甚至包含邮件用户信息。下面是 xfocus 的一个 Message-ID 的形式：

20040409085748.91B1.SAN@xfocus.org

其中就是由日期、时间、标识、邮件用户和 DNS 构成。

#### **Received:**

最可信赖的头。一般会有几条，形成站点列表，这些信息表明达到目的地过程中邮件所经过的服务器，该域都是邮件服务器自动插入的，spammer 可以伪造，但是在被伪造的那个点之后的是无法伪造的。这个列表从下往上表明了服务器路径，最上面的一条 Received: 是最终目的的系统或邮件服务器。每个邮件系统都有自己的 Received: 样式，因此，也可以通过该行来鉴别接收邮件的系统，比如：

```
Received: from xxx.com by bjmx5 (Coremail) with SMTP id
d6wPALnj10BHAMPI.2; Wed, 05 May 2004 02:41:14 +0800 (CST)
```

其中，xxx.com 就是发送邮件的机器的名称。这可能是被伪造的。而 id 可能可以帮助系统管理员来追踪垃圾邮件。一些系统会插入额外的信息，比如 IP 地址，如果 IP 地址和名称不能匹配就是被伪造了。如果，Received: 这一行是伪造的，那么 spammer 的插入点就应该是在上面一行。比如：

```
Received: from xxx.com (unknown [210.220.xxx.xxx])
by bjmx5 (Coremail) with SMTP id d6wPALnj10BHAMPI.2
for <spammer@xxx.net>; Wed, 05 May 2004 02:41:14 +0800 (CST)
```

### **3 邮件传输过程**

通常的邮件传递主要步骤由下面过程完成：

sender -> MUA -> MTA -> (routing) -> MTA -> MDA -> {filtering} ->  
MUA -> receiver

MUA: Mail User Agent, 邮件客户端程序, 比如 Foxmail、Eudora、OUTLOOK、  
mutt 等。

MTA: Mail Transport Agent or Message Transfer Agent, 消息传输代理。  
这部分程序负责存储和转发、发送 EMAIL。它从 MUA 或者其他的 MTA 接收到邮件后, 就存  
存在本地, 并分析收件人或者转发到其他的 MTA。在处理过程中, 它通常会编辑、添加邮件  
头内容。比如 Sendmail、Exchange 等。

MDA: Mail Delivery Agent, 邮件发送代理。这个程序负责将邮件发送给用户。  
MDA 通常处理某种特定发送操作。Unix 下典型的就是 /bin/mail, 它负责将邮件放到用  
户本地邮箱中。

MUA 通常编写的头部, 可以参考 RFC2822, 主要是:

From: # 发送者  
To: # 接收者  
Cc: # 抄送接收者  
Bcc: # 暗送接受者, 不能在邮件头中看到, 否者是不正常的  
Subject: # 主题  
Reply-To: # 回复时的接收者, 可以同 From:地址不一样  
Priority: # 优先级  
Resent-To: # 转发时用  
Resent-Cc: # 转发时用  
Date: # MUA 创建邮件时间

```
X-[something]:      # MUA 的个性化内容, 比如 X-Mailer: Microsoft
Outlook Express 6.00.2800.1158
```

在发送期间, 可以 MUA 还会插入一些附加头信息, 可以参考 RFC2045。

当 MTA 接收到邮件后, 可以插入的附加信息是:

```
From                # 本地邮件才可能添加, 或者是过滤器等增加的
Date                # 时间
Message-Id:         # 第一个 MTA 创建, 唯一的邮件标记
Received:           # 邮件路径
Return-Path:        # 表明怎么回到发送者
```

在 MTA 到 MTA 之间, 一般每个 MTA 都会添加 Received:, 从而形成一个 MTA 的列表, 能够用于分析邮件传递的路径。

最终 MTA 将邮件递交给 MDA, 可能添加的是:

```
Apparently-To:      # 如果没有 To: 的时候添加
From                # 本地邮件才参加
```

以上是通常情况下, 邮件头不断被插入新的域的过程, 但是, 正常的传递过程, MTA 等都不会对邮件头某个域的内容进行修改, 除非是发送者在 MUA 阶段就进行了修改 (如果没有控制 MTA 情况下)。对邮件来源的追踪, 主要分析的就是 received: 区域的内容。

## 4 邮件头分析实例

下面是一个真实邮件头的例子, 可以以此来为例分析 (其中仅仅修改了邮件地址和 IP

地址):

Return-Path: <[wu@xxx.com.cn](mailto:wu@xxx.com.cn)>

Delivered-To: [refdom@xfocus.org](mailto:refdom@xfocus.org)

Received: from mail.xxx.com.cn (unknown [211.167.xxx.xxx]) by xfocus.org (Postfix) with ESMTTP id 590F2160A9 for <[refdom@xfocus.org](mailto:refdom@xfocus.org)>; Thu, 6 May 2004 16:48:46 +0800 (CST)

Received: from mail.xxx.com.cn ([127.0.0.1]) by localhost (mail [127.0.0.1]) (amavisd-new, port 10024) with ESMTTP id 30543-01 for <[refdom@xfocus.org](mailto:refdom@xfocus.org)>; Thu, 6 May 2004 16:47:14 +0800 (CST)

Received: from risker.debian.org (unknown [218.18.xxx.xxx]) by mail.xxx.com.cn (Postfix) with ESMTTP id 32E0817DC17 for <[refdom@xfocus.org](mailto:refdom@xfocus.org)>; Thu, 6 May 2004 16:47:06 +0800 (CST)

Date: Wed, 5 May 2004 14:36:13 +0800

From: wlj <[wu@xxx.com.cn](mailto:wu@xxx.com.cn)>

To: [refdom@xfocus.org](mailto:refdom@xfocus.org)

Subject:

Message-Id: <[20040505143613.25dd214b.wu@xxx.com.cn](mailto:20040505143613.25dd214b.wu@xxx.com.cn)>

Mime-Version: 1.0

Content-Type: multipart/mixed;

X-Virus-Scanned: by amavisd-new at xxx.com.cn

上面的邮件头，用蓝色标记的是在 MUA 发送邮件时添加的头内容，其余的都是邮件经过 MTA 过程中添加的。

检查 Received: 的时候将从下向上分析。

*Received: from risker.debian.org (unknown [218.18.xxx.xxx]) by mail.xxx.com.cn (Postfix) with ESMTTP id 32E0817DC17 for <refdom@xfocus.org>; Thu, 6 May 2004 16:47:06 +0800 (CST)*

这是第一个 MTA 从 MUA 接收邮件时插入的头内容。MUA 的机器名是 risker.debian.org (这不是 MUA 的 DNS, 而只是他的机器名而已。), (*unknown [218.18.xxx.xxx]*)表示该机器的 IP 地址, 但是查询的 DNS 是 unknown 的。该邮件被 mail.xxx.com.cn 接收, 邮件服务器采用 Postfix, 而且采用的是 ESMTTP(扩展的 SMTP), 分配的 ESMTTP id 是 32E0817DC17, 传递目标是 [refdom@xfocus.org](mailto:refdom@xfocus.org), 接收时间为 Thu, 6 May 2004 16:47:06, 时区是+0800 (CST)。

接下来的 Received: 为:

*Received: from mail.xxx.com.cn ([127.0.0.1]) by localhost (mail [127.0.0.1]) (amavisd-new, port 10024) with ESMTTP id 30543-01 for <refdom@xfocus.org>; Thu, 6 May 2004 16:47:14 +0800 (CST)*

这是邮件服务器内部程序进行的一个处理过程, 因此 IP 地址为 127.0.0.1, 并且是 localhost 处理, (amavisd-new, port 10024) 表明这个处理程序是使用的 amavisd-new, amavisd-new 是一个用于邮件服务器的杀毒、过滤等的接口。

第三个处理 MTA 是:

*Received: from mail.xxx.com.cn (unknown [211.167.xxx.xxx]) by xfocus.org (Postfix) with ESMTTP id 590F2160A9 for <refdom@xfocus.org>; Thu, 6 May 2004 16:48:46 +0800 (CST)*

该过程表示邮件从服务器名为 mail.xxx.com.cn 传递出去, IP 地址为 211.167.xxx.xxx, 接收邮件的服务器是 xfocus.org, 采用 Postfix 服务程序, 也通常使用的 ESMTTP, 传递的目标是 [refdom@xfocus.org](mailto:refdom@xfocus.org), 日期为 Thu, 6 May 2004 16:48:46, 时区是+0800 (CST)

从这个真实例子可以看出，邮件的传递过程是：

- 1、risker.debian.org (MUA)
- 2、mail.xxx.com.cn (MTA) ——Thu, 6 May 2004 16:47:06
- 3、localhost (MTA 中的 amavisd-new) ——Thu, 6 May 2004 16:47:14
- 4、xfocus.org (MTA) ——Thu, 6 May 2004 16:48:46

整个过程经历了将近两分钟。

这是一个真实的正常邮件传递过程，但是实际上我们在追踪垃圾邮件过程中，这个传递过程中的 Received: 存在被篡改的可能，并且实际中也大量出现，因此，需要首先判断哪些信息是伪造的，哪些是真实的。

对于 Received:，最后的站点是接收者自己的邮件服务器，因此，最后的 Received: 是真实可靠的，除非自己的服务器已经不安全了。

## 5 邮件伪造

SMTP 协议在创建的时候并没有考虑到未来的邮件会成为垃圾，因此，安全性很差，邮件头可以任意创建、伪造和修改。

邮件头的伪造是很容易的，比如 Real Name, Return Address。在一些 MUA 上，比如 Outlook 也可以进行更改。一般情况下，发送者的很多内容不会被邮件服务器检查，服务器只关心接收者。

Spammer 当然需要进行伪造了。比如很多 ISP 都采用 TOS (Terms Of Service) 或者 AUP (Acceptable Use Policy) 来禁止那些非法使用。Spammer 通常用一些邮件程序将邮件转发到其他的邮件服务器，并且修改和伪造邮件头，避免被追踪或者被 ISP 处理。

追踪邮件来源的关键就是识别伪造内容，并获得真实信息，根据真实信息进行查询。对于 Received: 域来说，一般情况下，我们从下面的痕迹出发：

- 时区出错。比如 -0600 (EST)，这就是错误的，EST 是在 -0500。
- 时间误差，通常情况下，一个邮件传递过程不会太长，假如这个邮件传递经过了几天

甚至更长时间，那么就可能是被伪造的（除非是传递过程中出错）。

- IP 地址错误，比如出现 xxx.xxx.xxx.0 这样的地址。

## 6 垃圾邮件分析

### 实例分析一

这是一封真实的商业广告垃圾邮件，邮件内容是介绍一个婴儿用品公司的网站。当然，邮件内容中有联系人、联系电话、联系 email、邮编地址等东西，追查就很直接，但是本文引用该邮件主要是简单分析邮件头内容。

邮件头内容如下：

```
Return-Path: <dfd@dfd.com>
Delivered-To: refdom@xfocus.org
Received: from dfd.com (unknown [221.232.11.40])
    by xfocus.org (Postfix) with ESMTP id 399521C124
    for <refdom@xfocus.org>; Mon, 24 May 2004 11:07:41 +0800 (CST)
From: "bbcsc" <dfd@dfd.com>
Subject: =?GB2312?B?0KGxvrS0tPPStcrmtvmxptXQycw=?=
To: refdom@xfocus.org
Content-Type: multipart/mixed;

boundary="=_NextPart_2rfkindysadvnqw3nerasdf";charset="GB2312"
MIME-Version: 1.0
Reply-To: dfxc@vip.sina.com
Date: Mon, 24 May 2004 11:07:45 +0800
X-Priority: 3
Message-Id: <20040524030745.399521C124@xfocus.org>
```

现在来对该邮件进行简单的分析。首先看

```
Received: from dfd.com (unknown [221.232.11.40])  
by xfocus.org (Postfix) with ESMTTP id 399521C124  
for <refdom@xfocus.org>; Mon, 24 May 2004 11:07:41 +0800 (CST)
```

我的邮件服务器 xfocus.org 当然是可信的,因此这一条 received 信息也是可靠的,只是可能其中的一些内容并不是真实可靠的。邮件来自一个机器名为 dfd.com 的,IP 地址为 221.232.11.40,邮件接收时间是 Mon, 24 May 2004 11:07:41 +0800 (CST)

简单的以 dfd.com 检查,得出 IP 地址为 dfd.com [209.35.182.41],很容易可以知道这个 dfd.com 只是一个名字而已。

IP 地址 221.232.11.40 是属于湖北的。

该邮件的发送者是 From: "bbcsc" <[dfd@dfd.com](mailto:dfd@dfd.com)>, 而回复地址是: Reply-To: [dfxc@vip.sina.com](mailto:dfxc@vip.sina.com)

实际上, [dfd@dfd.com](mailto:dfd@dfd.com) 是伪造的了,但是回复地址却可能是真实的。

这种商业垃圾邮件一定是使用了一些垃圾邮件发送工具,能够伪造发送者地址、机器名,并且可以直接传递邮件,比如自带 SMTP。

## 垃圾邮件实例二

下面是我接收到的一个真实的垃圾邮件(内容是什么美国之音)头信息,比上一封的分析将显得复杂一些,但也并不困难:

```
Received: from imu.edu.cn (unknown [221.192.19.3])  
by bjmx1 (Coremail) with SMTP id +nQSADJRnkAIABMD.2  
for <nyyu@tom.com>; Sun, 09 May 2004 23:42:06 +0800 (CST)
```

X-Originating-IP: [221.192.19.3]  
Received: from [218.105.225.91] uid 9369 星期日, 09 五月 2004  
07:55:59  
Message-ID: <20040509179343.1649@mail.14m18.com>  
From: [honglove@imu.edu.cn](mailto:honglove@imu.edu.cn)  
To: [nyyu@tom.com](mailto:nyyu@tom.com)  
Date: Sun, 9 May 2004 08:43:21 -0700  
Subject: American English , 70  
Mime-Version: 1.0

这个邮件接收对象是 [nyyu@tom.com](mailto:nyyu@tom.com), 这并不是我的邮件地址, 很明显, 我的邮件地址被放在了 Bcc 里面了, 所以在邮件头中是没有显示的。

可以肯定的是, 这个是真的:

Received: from imu.edu.cn (unknown [221.192.19.3])  
by bjmx1 (Coremail) with SMTP id +nQSADJRnkAIABMD.2  
for <[nyyu@tom.com](mailto:nyyu@tom.com)>; Sun, 09 May 2004 23:42:06 +0800 (CST)

我们假设整个邮件都没有被伪造过, 然后来分析, 可以看出一些有趣的东西:

该邮件的传递过程是:

- 1、218.105.225.91 (MUA) — Sun, 9 May 2004 08:43:21 -0700
- 2、[mail.14m18.com](mailto:mail.14m18.com) (MTA)  
或 [imu.edu.cn](mailto:imu.edu.cn) (unknown [221.192.19.3])  
— uid 9369 星期日, 09 五月 2004 07:55:59
- 3、[bjmx1](mailto:bjmx1) (FROM: 221.192.19.3) — Sun, 09 May 2004 23:42:06 +0800 (CST)

- 按照邮件传递时间的分析:

邮件的传递时间是很有意思的，在步骤 1 中，时间是 **08:43:21 (时区为-0700, 美国山区时区)**，而到第 2 步，时间变成了 **07:55:59**，最后接收时间是 **23:42:06 (时区为+0800)**，如果统一时区，那么接收时间是 **08:42:06 (时区为-0700)**。从这里发现步骤 1 和 3 时间有错位，接收时间比邮件创建的时候更早，难道邮件进入了时间隧道？合理的解释是下面的一种可能：

- 步骤 1 和 3 中的系统时间不是标准时间，服务器出现了一些误差（时间误差不多，很有可能）；
- 步骤 1 中邮件头是伪造的；

但是步骤 2 的时间（**07:55:59**）就有问题了，时间相差太多。但是因为没有显示时区信息，那么有可能是：

- 邮件头是伪造的；
- 所处时区不一样，推测步骤 2 的 MTA 位于时区为-0600，也就是美国中部时区，调整为-0700 时区的话，时间就是 **08:55:59** 了（从时间上看很有可能）；

如果以从这样分析，我们可以得出一个结论：邮件是从地理位置的美国偏中部，传递到美国偏西部，然后再传递到中国的。

- **按照邮件传递地址的分析：**

传递路径中，第 3 条信息是我的邮件服务器的内容，因此可以信任该内容是完全真实的，并且它接收的邮件来自机器名（不是域名）为 `imu.edu.cn (unknown [221.192.19.3])`，但是 DNS 并没有得到，是 `unknown`。

而第 2 步骤就非常有意思了，我在这里设置了一个或者关系，因为，假设如果每个 MTA 都插入 `Received` 的内容话，那么第 2 步就应该不存在或者关系了，原因是，从 3 可以得出结论，前一 MTA 应该是机器名为 `imu.edu.cn[221.192.19.3]` 的系统发出的，但是，

我们从 Message-ID:得出的结论应该是: 这个 MTA 为 *mail.14m18.com*, (应该不会存在域名为 *14m16.com* 的机器名为 *imu.edu.cn* 的) 但是, whois 查询的结果告诉我们, *14m18.com* 是根本不存在的, 也就是说, Message-ID 是完全属于伪造, 并且 *221.192.19.3* 地址经过查询, 属于河北省的, 假设 *imu.edu.cn* 是一个正确的域名 (来自这个域名的邮箱), 查询域名地址为 *202.207.0.x* 范围, 由此可以判断邮件发送者地址是伪造的。

经过时间和地址的两种办法分析之后, 得出的结论只有:

邮件是通过暗送传递出来并到达我的邮箱, 邮件传递过程被伪装或篡改了, 根本不可信任, 但是, 邮件来自河北省 IP 地址为 *221.192.19.3* 这是可以肯定的 (但是现在我没有办法与该主机通讯, 无法察看该主机是否是被利用的邮件服务器)。

### 垃圾邮件实例三

这是一封病毒 (W32/Lovgate.x@MM) 邮件:

```
Return-Path: <envoywaltsimer@diplomats.com>
Delivered-To: virus-quarantine
X-Envelope-To: <refdom@xfocus.org>
X-Quarantine-id: <virus-20040513-125224-16799-09>
Received: from diplomats.com (unknown [202.195.227.201])
    by xfocus.org (Postfix) with ESMTP id 381FD18EC5
    for <refdom@xfocus.org>; Thu, 13 May 2004 12:51:44 +0800 (CST)
From: envoywaltsimer@diplomats.com
To: refdom@xfocus.org
Subject: Hello
Date: Thu, 13 May 2004 12:54:13 +0800
MIME-Version: 1.0
```

```
Content-Type: multipart/mixed;  
    boundary="-----_NextPart_000_0008_927B3592.2CE9A115"  
X-Priority: 3  
X-MSMail-Priority: Normal  
Message-Id: <20040513045145.381FD18EC5@xfocus.org>  
X-AMaViS-Alert:    INFECTED,    message    contains    virus:  
W32/Lovgate.x@MM
```

邮件内容应该都是真实的，只是其中一些部分被邮件服务器上的反病毒插件作了修改。病毒感染者是 [envoywaltsimer@diplomats.com](mailto:envoywaltsimer@diplomats.com)，虽然邮件内容和标题都可能做得很有欺骗性，但是伪造邮件头的病毒还很少。

## 7 总结

从前面的几个实例简单分析过程，我们可以知道一般有下面内容能够给我们提供需要的线索来追踪邮件来源：

- 邮件内容。特别是一些商业垃圾邮件，一般内容中提到的联系方式、网站等是可信的。
- Reply-to，有的时候伪装了发送者地址，但是发送者可能希望得到回复，所以 reply-to 可能就提供的是真实的邮件地址。
- 最终邮件服务器的 received: 的内容，除非你的服务器被控制了。注意，不要相信其中的机器名，这些机器名通常都修改为某个域名一样的。但是它的 IP 地址你是可以相信的，虽然用来追踪可能显得还很弱。

从上面的分析也很容易看出，通过对邮件的分析，我们一般能够找到可能接近源头的某个邮件地址或者一个 IP 地址（这个 IP 地址可能是一个受害者），用这些信息来追查，依然存在很多难度，毕竟有些事情不是某个人可以完成的，但是却在某些特殊应用方面能够提供

不小的帮助。

## 8 参考

- 1、IP 地址查询: <http://www.apnic.net/apnic-bin/whois.pl/>
- 2、Whois: <http://www.whois.org/>
- 3、<http://www.fags.org/fags/>