IBM 研究

# 僵尸吸干邮件服务器的生命

(新进展 "摘自 LISA 2010 演示文稿)

**Wietse Venema**

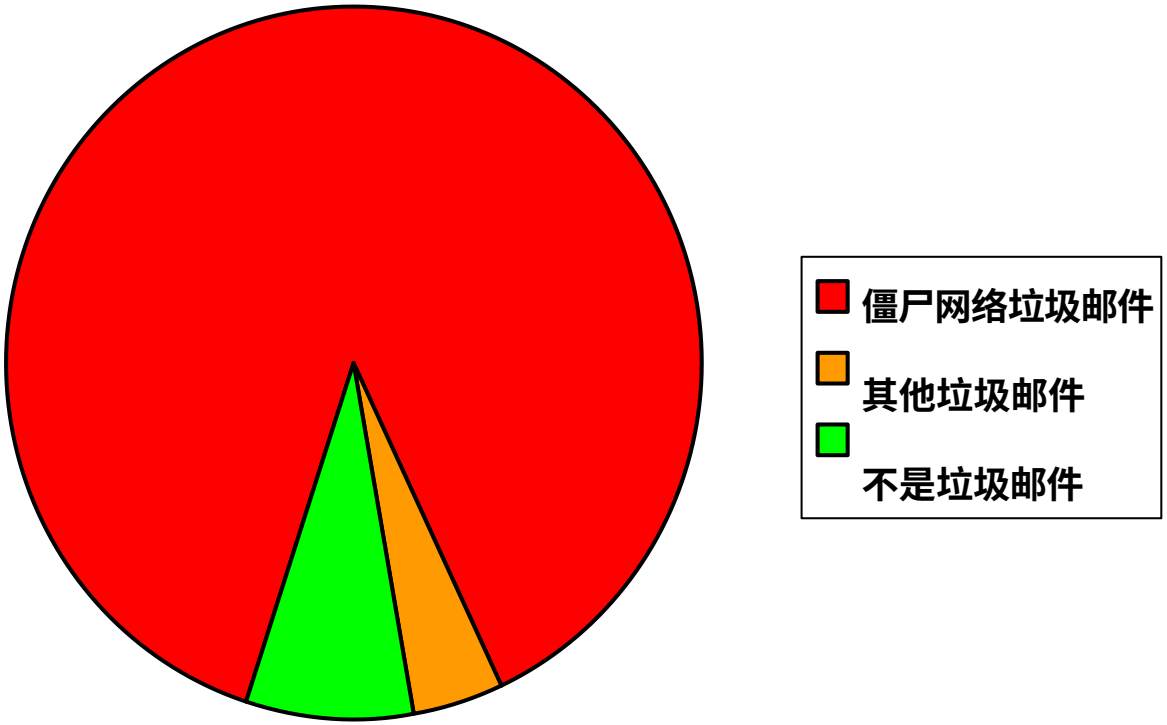
**美国纽约州霍索恩 IBM T. J. 沃森研究中心**

# 不断变化的威胁

- 2009:您建立的邮件系统具有世界一流的电子邮件发送性能。
  - *问题*：你的世界级邮件系统现在大部分资源都用在了*无法投递邮件上*。
    - *解决方案*：更聪明地工作。

不断变化的威胁

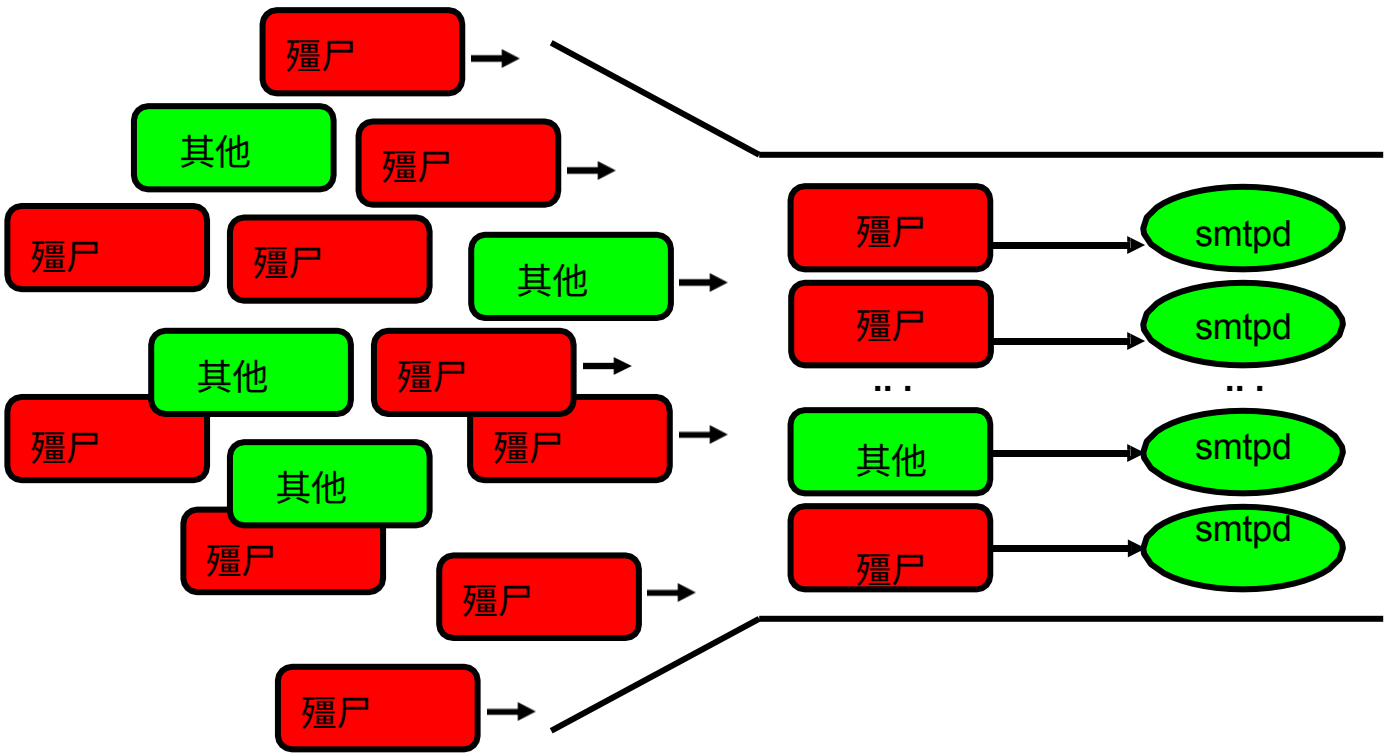
# 92% 的邮件是垃圾邮件，95% 的垃圾邮件来自僵尸网络




资料来源：MessageLabs Intelligence 报告，2010 年 8 月：来源：MessageLabs Intelligence 报告，2010 年 8 月

不断变化的威胁

# 僵尸使邮件服务器端口繁忙




不断变化的威胁

# 僵尸吸干邮件服务器的生命

- 最坏的例子风暴僵尸网络

13:01:36 postfix/smtpd: connect from [x.x.x.x].

13:01:37 postfix/smtpd: reject：RCPT from [x.x.x.x]：550 5.7.1 等等等等

**13:06:37 postfix/smtpd：来自 [x.x.x.x.x] 的 RCPT 后超时。**

– RFC 5321 建议服务器端超时 5 分钟。

- Postfix 根据标准实现 SMTP。

  - 结果：所有 SMTP 服务器端口都被风暴僵尸占用。

不断变化的威胁

# 邮件服务器超载策略

## 主要针对中小型网站

- 假设：僵尸问题在改善之前（如果有的话）会变得更糟。
- 临时超负荷：
  - 工作速度更快：每个 SMTP 客户端所需的时间更短（负载分流）。
- 持续超负荷：
  - 更努力地工作：处理更多 SMTP 客户端（叉车解决方案）。
  - 更聪明地工作：在上游阻止垃圾邮件机器人（屏幕后）。

# 临时超负荷战略

- 工作速度更快：每个 SMTP 客户端花费的时间更少。
  - 减少时间限制、被拒绝指令的数量等。
    - 只需 21 行代码即可实现自动配置切换（2007 年）。
  - 会耽误*一些*合法电子邮件的发送。
    - 来自网络延迟或数据包丢失较大的站点。
    - 来自邮件列表，超时严重。
  - 收到*一些*合法邮件总比*没有邮件好*。
    - 只要过载情况是暂时的，就没问题。

不断变化的威胁

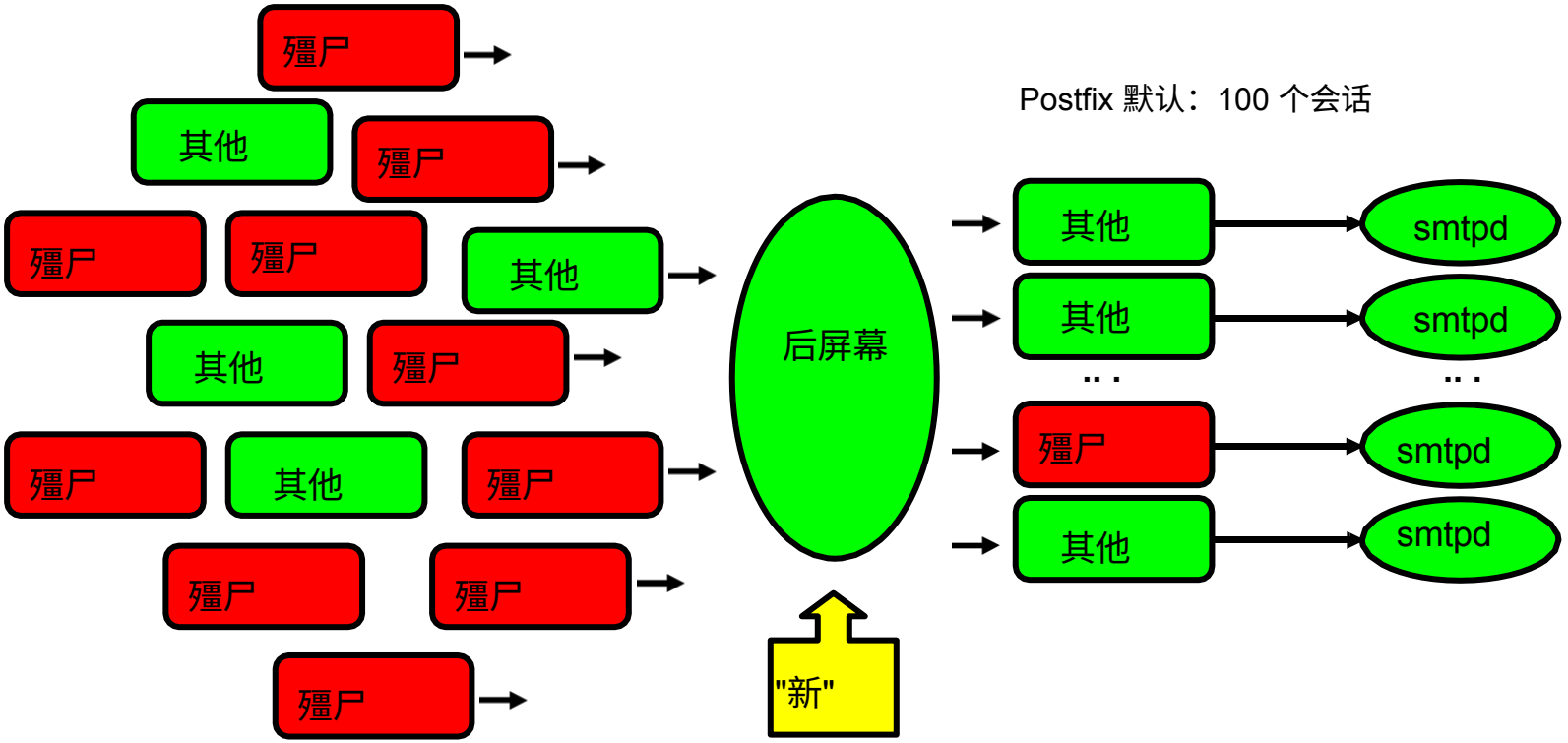
# 持续超载策略

- 努力工作：配置更多邮件服务器进程。
  - 蛮力叉车法。
  - 如果你能负担得起网络、内存、磁盘和 CPU，那就没问题。
- 更聪明地工作：让僵尸远离服务器。
  - 服务器连接前过滤器。
  - 为合法电子邮件保留更多 SMTP 流程。

不断变化的威胁

# 持续过载 - before-smtpd 连接过滤器

## 以前的工作OpenBSD spamd、MailChannels TrafficControl、M.Tokarev




不断变化的威胁

# 筛后(8) 挑战与机遇

- 僵尸会在几小时内被列入黑名单[1]。
  - 机会：拒绝匆忙发送邮件的客户。
    - 语速过快的客户端：预问候、命令流水线。
    - 其他公然违反协议的行为。
    - 陌生人连接时出现假的 "临时 "错误（灰名单）。
- 僵尸避免重复向同一网站垃圾邮件。
  - 挑战：为单个连接确定 "是僵尸"。
    - 使用 DNS 白名单和黑名单作为共享情报源。

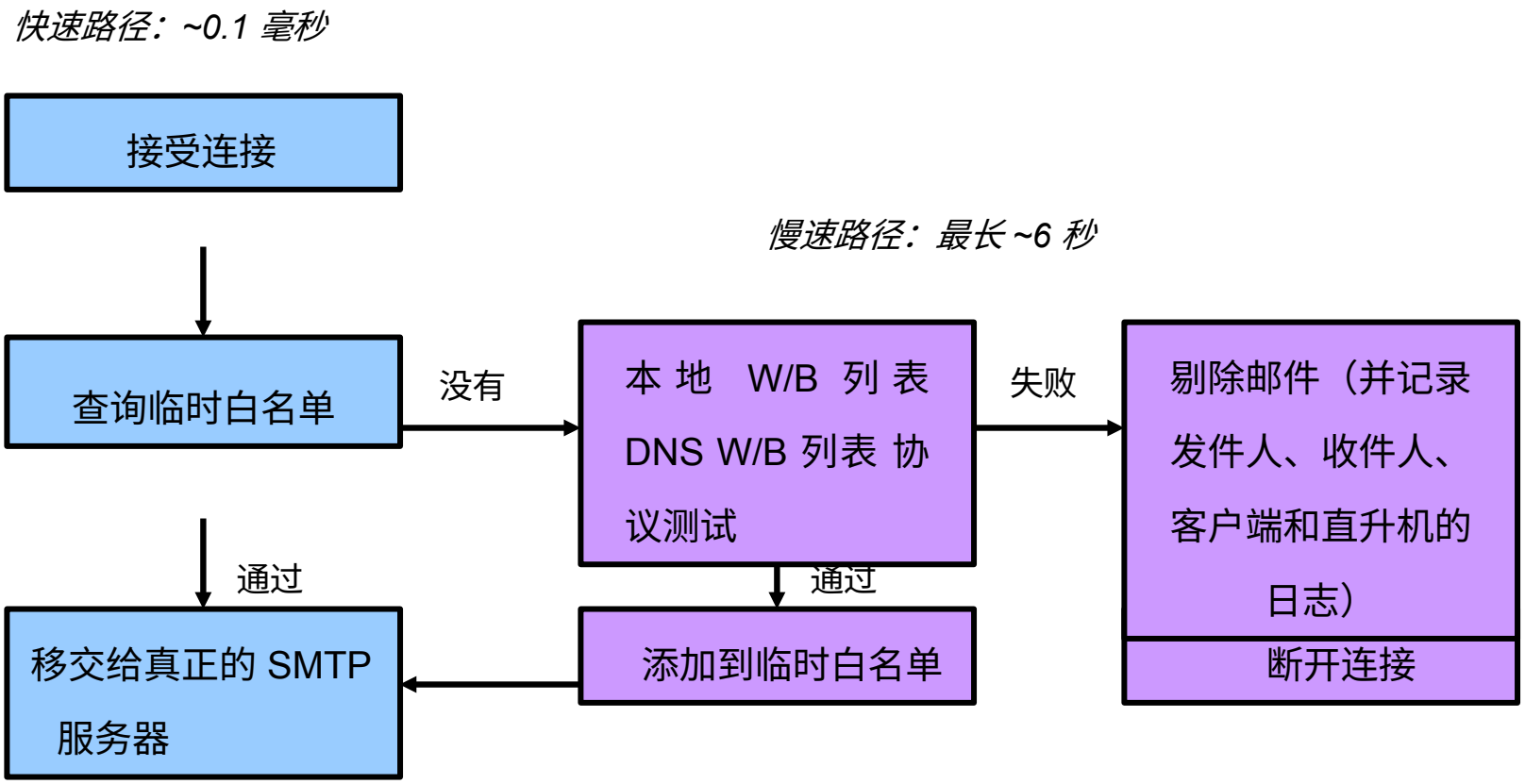
[1]Chris Kanich 等，Spamalytics：垃圾邮件营销转化的实证分析》，CCS 2008。

# 电子邮件等 DNS 白名单和黑名单

- 最初由 ISC 的 Paul Vixie 设计。
  - 互联网软件联盟提供 DNS、DHCP 等的参考实现。
  - 要了解 *mail.abuse.org* 是否列出了 *1.2.3.4* 地址，请查询 *4.3.2.1.mail.abuse.org* 的 IP 地址。
- 常用提供商：spamhaus.org、spamcop.net、barracudacentral.org。
  - 垃圾邮件陷阱和其他传感器。
  - 一些 DNS[BW]L 提供商对小用户免费。

# postcreen(8) 工作流程

## 一个守护进程同时筛选多个连接




不断变化的威胁

# 检测提前发言的垃圾邮件机器人（问候语）

- 良好的 SMTP 客户端会等待 SMTP 服务器的问候语：

*SMTP 服务器：* **220 server.example.com ESMTP Postfix<CR><LF>**

*SMTP 客户端：* **EHLO client.example.org<CR><LF>**

- Sendmail *greet_pause* 方法：在发送 220 个问候语之前等待几秒钟。
  - 很少有客户过早打招呼。
  - 更多的客户几秒钟后就放弃了。
  - 手动白名单。

不断变化的威胁

# 向捕狗员提出的问题

- 问：如何快速了解房屋是否养狗?
- 答：按门铃，狗立即吠叫。

- postscreen(8) 对僵尸网络僵尸使用了类似的技巧。

不断变化的威胁

# 让僵尸狂吠--多行问候语陷阱

- 好的客户会等待完整的多行服务器问候语：

*邮件服务器：* **220-server.example.com ESMTP Postfix<CR><LF>**

*邮件服务器：* **220 server.example.com ESMTP Postfix<CR><LF>**

*良好客户端：* **HELO client.example.org<CR><LF>**

- 许多垃圾邮件发送者会在多行服务器问候语的第一行之后立即说话：

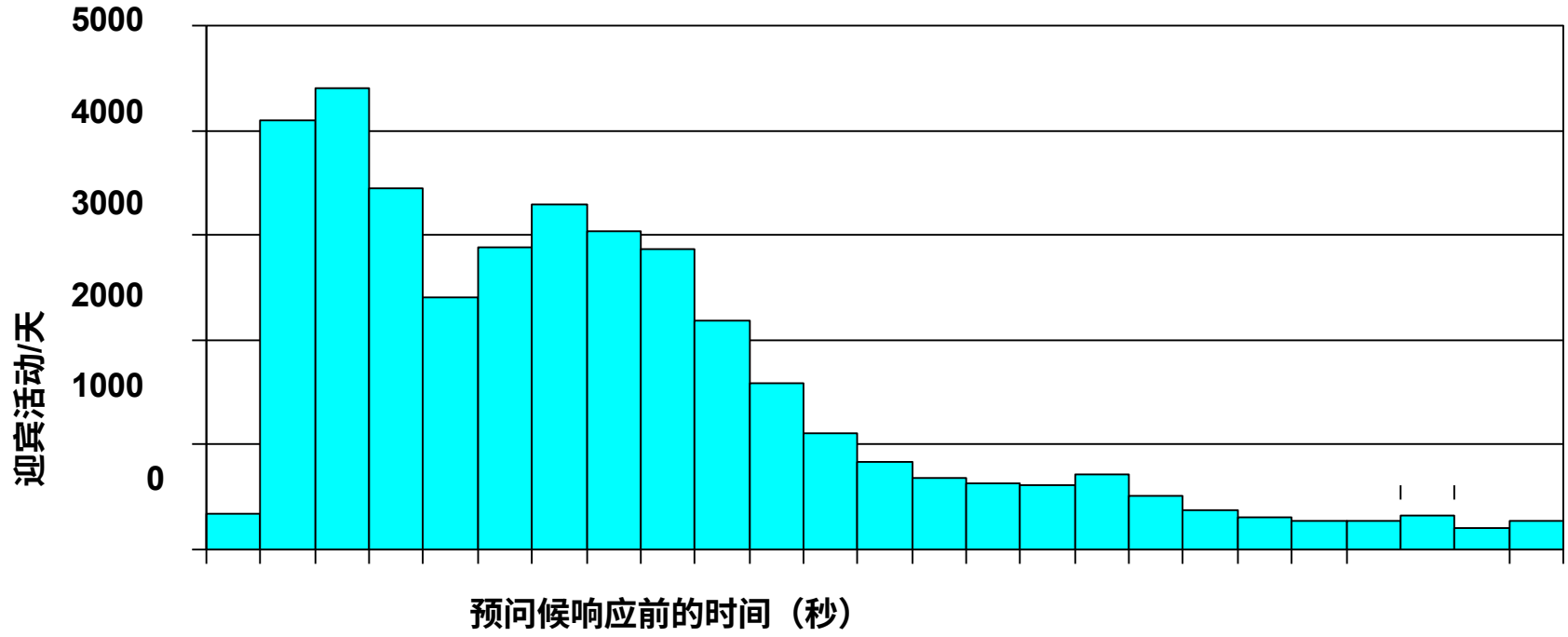
*postscreen：* **220-server.example.com ESMTP Postfix<CR><LF>**

*垃圾邮件机器人* **HELO i-am-a-bot<CR><LF>**

不断变化的威胁

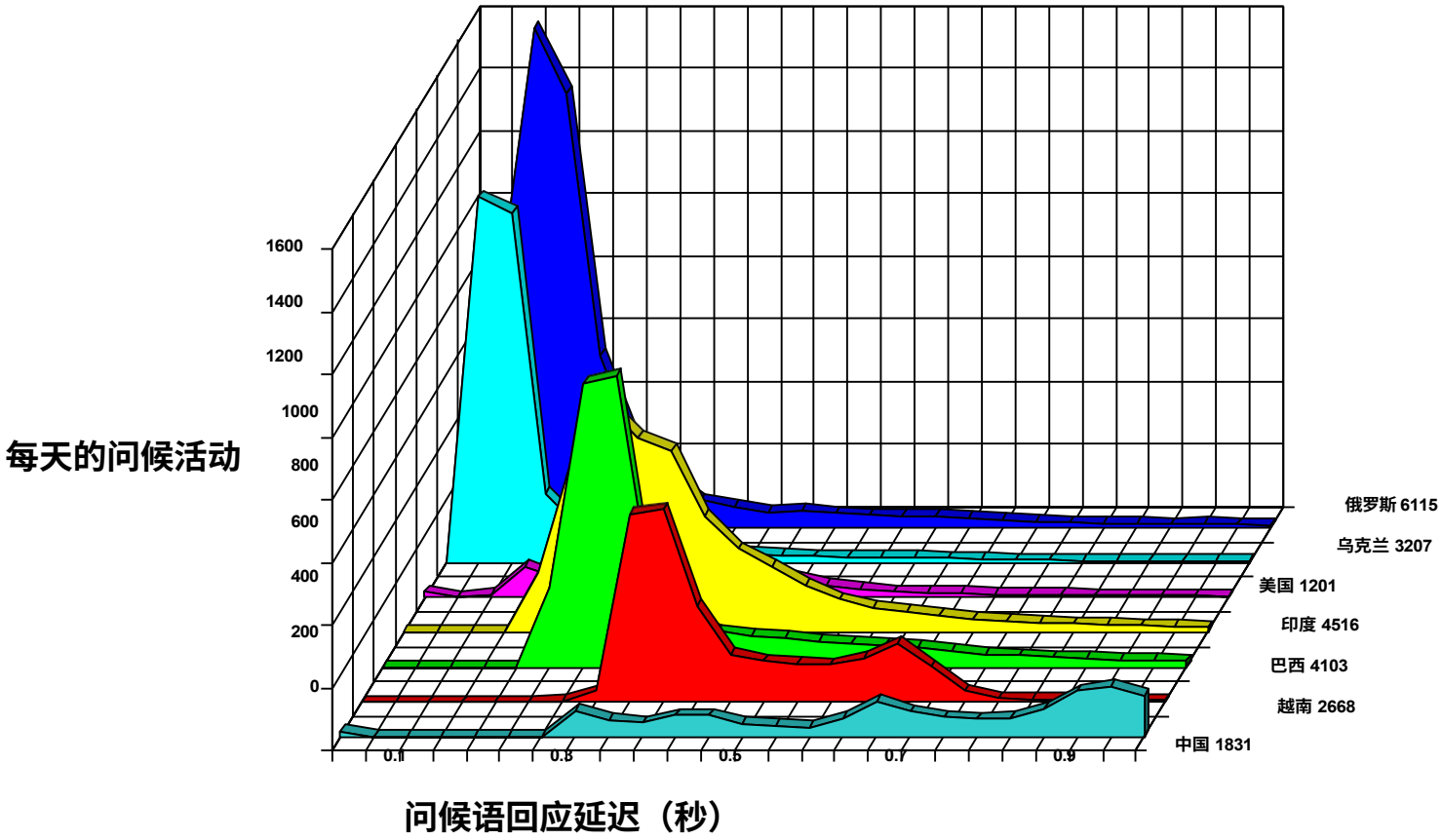
# 超过 60% 的机器人在 mail.charite.de 打招呼

## 8% 未列入 DNS 黑名单。柏林，2010 年 8 月 26 日 - 9 月 29 日



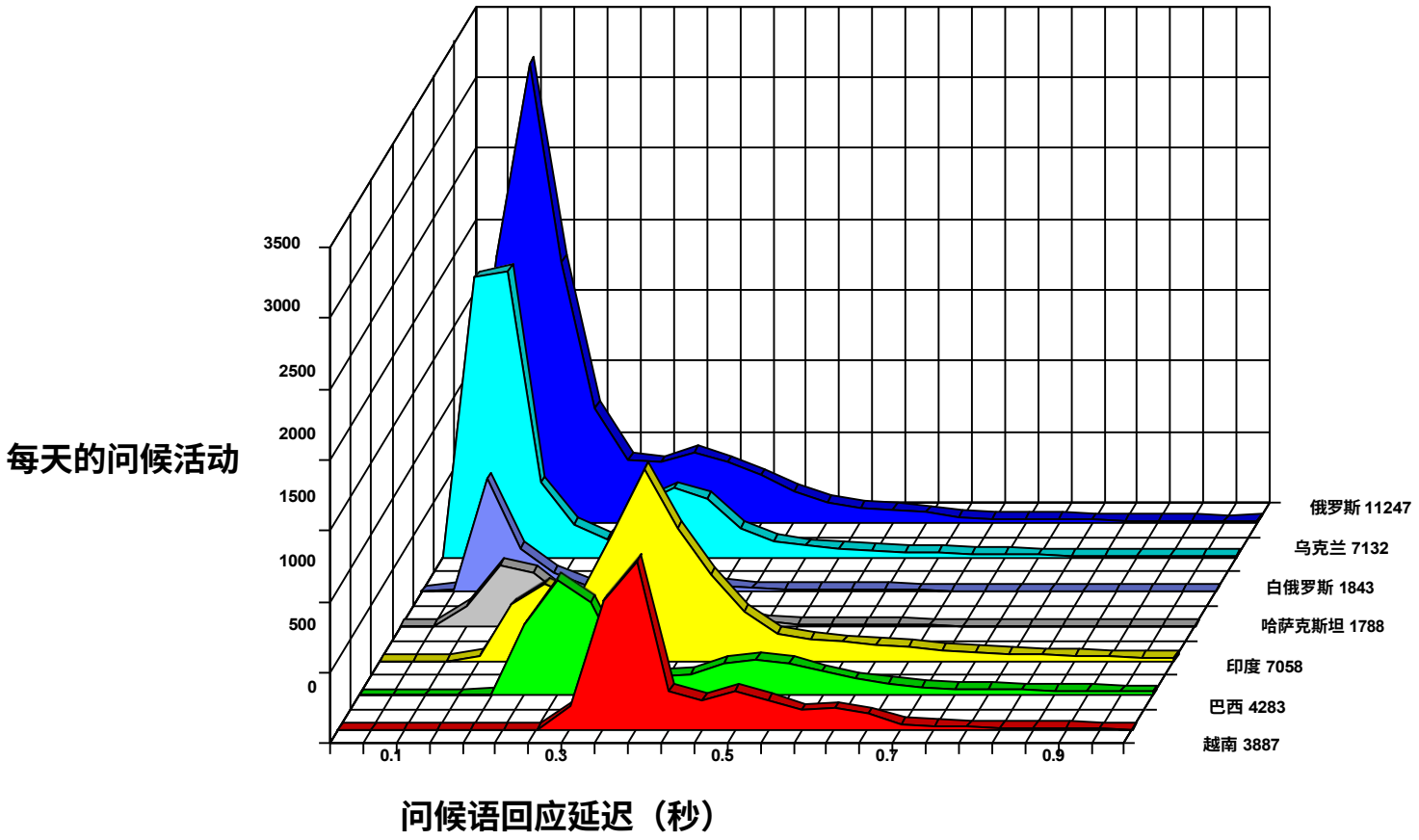
# 超过 60% 的机器人在 mail.charite.de 打招呼

## 8% 未列入 DNS 黑名单。柏林，2010 年 8 月 26 日 - 9 月 29 日



# 超过 70% 的机器人在 mail.python.org 打招呼

## 1% 未列入 DNS 黑名单。阿姆斯特丹，2010 年 9 月 16 - 29 日



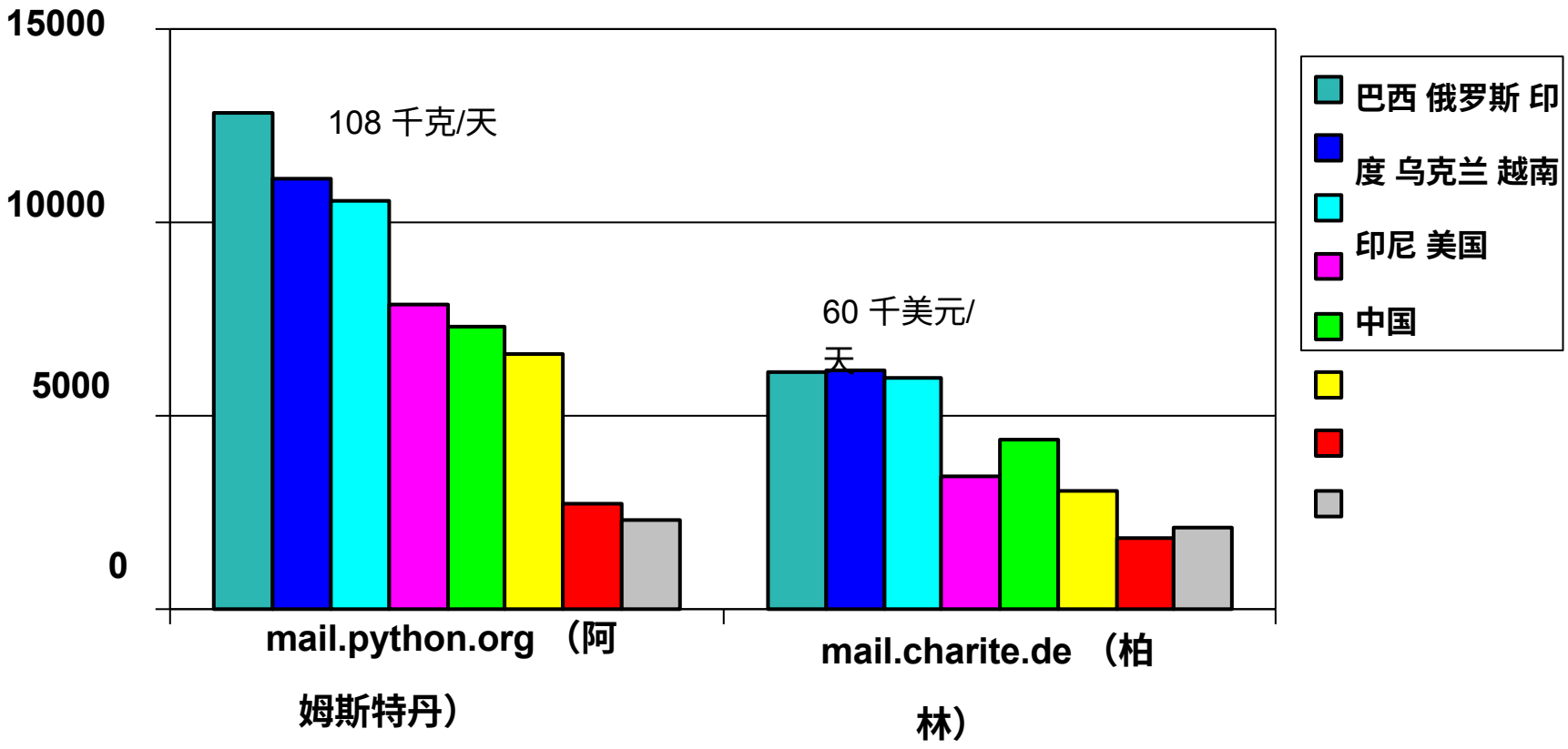
# 垃圾邮件负荷因接收者和时间而异

- 不同接收器的垃圾邮件负荷：
  - 当今大多数垃圾邮件都是由少数几个国家发送的，但不同的接收方看到的发送量是不同的。
- 每天不同时段的垃圾邮件负荷量：
  - 垃圾邮件是 24 小时运作的，但垃圾邮件机器人不是。
    - SPAM 的发送时间往往比 HAM[1] 晚。

(1) S. Hao 等，用 SNARE 检测垃圾邮件发送者：时空网络级自动信誉引擎》。Usenix Security 2009。

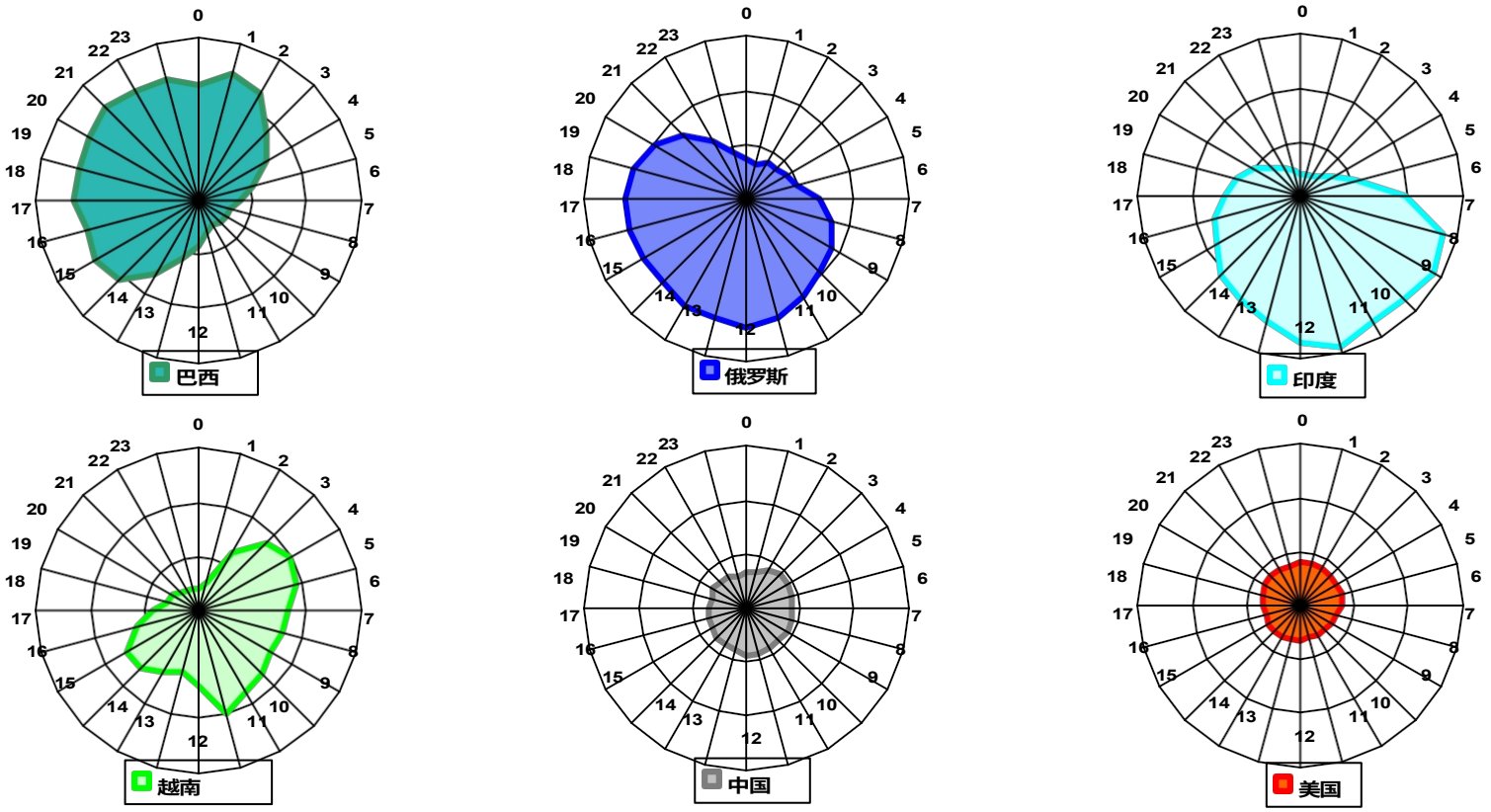
# 欧洲小型网站每天的垃圾邮件连接数

根据 zen.spamhaus.org 统计的垃圾邮件，2010 年 9 月 3 - 23 日




不断变化的威胁

根据 zen.spamhaus.org 提供的数据，2010 年 8 月 26 日至 9 月 29 日按来源国和时间分列的 mail.charite.de UTC+2 垃圾邮件数量




不断变化的威胁

# Postcreen(8) 结果和状态

- 并行、加权 DNS 白名单/黑名单查询。
- 静态白名单/黑名单，动态 "快速路径 "缓存。
- 试点结果（小型站点，每天最多连接 20 万次）：
  - 预先问候（说话太早）：多达 ~10% 未列入 DNS 黑名单。
  - 流水线作业（多命令）：~1% 的垃圾邮件机器人。
  - 僵尸挂起（读取超时）：~1% 的垃圾邮件机器人。
- 随着僵尸网络的发展，还将增加其他协议测试。
- 开始规划扩展接口。

不断变化的威胁

- 预计 2011 年初随 Postfix 2.8 一起发布。

不断变化的威胁