Comment #: 8

Comments for the FTC & NIST Email Authentication Summit

(Matter Number P044411)

Hadmut Danisch

September 27, 2004

Background and summary

FTC and NIST publicly announced an Email Authentication Summit to be held in Washington DC November 9-10, 2004, and requested comments on a list of 30 questions or any other issues in connection of email authentication.

I am the author of the RMX drafts[1], first published in 2002. The RMX drafts initiated the discussion about email authentication as a spam protection and thus the subject of this summit. The well-publicized proposals the Request for Comments mentions, SPF, CallerID, and SenderID were based on and directly derived from RMX. I am working on communication, network, and email security since the mid-1990's and participated in the IRTF's and IETF's anti spam efforts from the very beginning.

This paper comments on some of the questions posed by FTC and NIST and points out why email authentication alone will not solve the spam problem. It therefore proposes a simple technical measure which could – in conjunction with legal measures and the authentication and authorization discussed on this summit – effectively reduce the amount of spam.

Table of Contents

[1] see http://www.danisch.de/work/security/antispam.html

Authentication does not protect against spam

Email authentication and authorization **do not protect against spam**. They do protect against forgery. Since spam is usually sent as forged email, the protection against spam is a lucky side effect.

Until now, spam emails were usually forged, because spammers had absolutely no reason to not forge, but good reasons to do so. But since authentication methods evolve, spammers begin to stop forging and to use their own domains. Domain administrators found that most incoming messages authorized by SPF records are in fact spam messages. Ironically, some spam filters gave those spam messages a better score than plain email because of those SPF records.

Email authentication/sender authorization is only the first step. It allows to identify the spammer, but then further measures are required. Unfortunately, these measures are more complicated and more difficult than the first authentication/authorizaton step. If you want to defend against spam, then you need to solve the problem of how to proceed once you've identified the sender of spam.

There are currently different proposals on how to cope with that, but it is just the most publicized proposals SPF and SenderID, which completely ignore that problem. They are designed primarily for marketing and patenting purposes and not as technical security solutions, and thus are limited to the easy part of the problem only.

A former proposal of myself was to query the whois entry of the sender's domain, to identify the person responsible for the domain, and to reject e-mail if the whois database does not tell such a person. A third step would be to black- and whitelist and maybe to prosecute those responsible persons. Other proposals use domain-based reputation and accreditation systems.

Both approaches do have a significant flaw: They do not work on a world wide scale. They might be appropriate in one country, but not in a different one, and could be even unlawful. For example, the approach based on the whois database or the impressum given by a domain might be working in Germany, but not acceptable or unfeasible in the United States. On the other hand, reputation databases might be working in the U.S., but could violate german privacy laws. And both methods will not work in countries not under rules of law or which allow spamming.

You will not find a method that will work for the whole world. You will need per-country solutions. That's what makes things difficult, because spam is a cross-border problem.

To cope with this problem, I propose the method described below.

Proposed solution: .com considered harmful

It will be impossible to find a general method of how to proceed after indentifying the sender of spam by authentication. The legislations, the mentalities, the ideas of identity, and even the will to prevent spam are far too different all over the world. The unpleasant consequence is that every country on the world will have its own invidual way to cope with spam (or not), depending on the nature of the local legislation. In some countries spam might be even lawful. Other countries might not even have a working power and legislation. States which tolerate drug dealers, give shelter to terrorists, suffer from corruption, or have civil wars cannot be expected to prosecute spammers.

Therefore, it must be the recipient's decision, from which legislations and countries he wishes to accept, deny, or treat email differently. Mail Transfer Agents (MTA) should be configurable in from which countries they accept and how they treat email. Email from a country known to effectively prevent spam could be accepted always. Email from a country with not so good prevention could undergo further tests. And email from countries with no spam prevention might be completely blocked – depending on the MTA administrators and recipients personal taste and preferences. Everyone could decide on his own whether to take email from spam-friendly countries or to require better authentication. You can compare it to the Visa Waiver Program for US immigration – some countries are trusted more than others. The idea is to do the same with email, except for the fact that every MTA operator or domain owner would be able to decide for himself which countries are trusted more or less.

A major problem are the generic top level domains, especially `.com`. There is no reliable way to identify the country of the owner of the email sender's domain, since the whois databases are incomplete, inconsistent, incorrect, to slow and undersized, and suffer from a large number of different formats and languages. It will not be possible to ensure the correctness of the whois database for `.com` and the other gTLDs.

I therefore propose to not use generic top level domains for sending email anymore and to allow MTAs to block all such email. This might appear to be absurd at a first glance, but it is simple, feasible, effective, and cheap. If a company has the domain `example.com`, it may still continue to have a website and to *receive* email for that domain, e.g. for `support@example.com`. But for sending email it would have to use a domain like `example.us` or `example.com.us`, as already usual in many countries around the world. Obviously, such a `.us` domain must be given only to persons and organizations who reside in the United States and underly US laws and US jurisdiction. Any `.com` domain owner outside the USA must apply for a domain under the ccTLD of his own country for sending email. They could still use their `.com` domain for web sites and for *receiving* email. So the gTLDs will still be in use and available for international business, but they are limited to passive services, like WWW and receiving email. See my comment to Question 25 (page 10) for the impact on phishing.

The receiving MTA can now be easily configured to treat email differently depending on the country code, which could easily be seen from the sender's domain after authentication: The good ones pass, the bad ones are rejected, and the questionable ones undergo further checks or go in a special mail folder. Spam from 'good' countries will be kept low because the sender can be identified and prosecuted (what constitutes a country to be a 'good' one).

As a consequence, the internet will (temporarily) be split into the good, the bad and some gray countries, depending on their ability and readiness to prevent spam. Every country that wants to participate in the world wide email network must do its job to prevent spam. Since email has become essential for business, the industries and the governments of states will be interested to stop spam. Companies and private users in spam-friendly countries might still find ISPs in better countries willing to be their guarantor and to relay their email under their domain, maybe after applying countermeasures like filters or traffic limitations.

Fraud and Phishing not limited to Email

It is too shortsighted to focus on email only. Email might currently be the main battlefield for spam, fraud, and phishing, but it would be disastrous to limit the scope to email and to run into the same problem again soon.

There can be no doubt that spammers will find new internet services to abuse, e.g. chats, instant messaging, news. Voice over IP will soon become a standard application. Be prepared to be flooded with machine generated mass phone calls in the style of spam and phone fraud in the style of phishing mails, once Voice over IP becomes a standard application for cheap phone calls.

Any authentication method to be deployed on a world wide scale should be usable for other services as well and not limited to email only.

Correction to the Request for Comments - Capabilites of RMX and RMX++

The Request for Comments states that current proposals are limited to authenticate the domain part of a sender address only, i.e. if a message claimed to be from `abc@ftc.gov` the private market authentication proposals would authenticate that the message came from the domain `ftc.gov`, but not the particular email address `abc` at this domain.

This is incorrect. The RMX draft describes methods to authenticate the full email address.

The RMX++ proposal[2] goes further and allows to base the authorization on more message properties than just the sender's address. It allows to also consider properties like the transmission time, the number of messages sent, the subject, the message id, the recipient's address, the receiving MTA's address, even the country the receiving MTA resides in. It allows to dynamically generate the authorization records and can even hide the authorization check completely in the domain owner's computers, thus allowing any arbitrary method of authentication. It allows the domain owner to account every message sent and to detect abnormal behavior of sending machine, e.g. when they were hijacked by hackers.

Current situation of anti spam development

The current situation is chaotic. The IETF and IRTF have failed to form a standard, after wasting months of time with more or less useless discussions, trying to please commercial interests, being focussed on marketing and the press instead of proper design and development of a security protocol. The MARID working group was just

[2] now included in the RMX draft

dissolved after failing to fulfill its goals. The IETF and IRTF have failed to defend themselfes against malicious and harmful influence by commercial parties. They spoiled the chance to have an independent and free standard and delayed the development for almost two years.

RMX was first published in 2002. Now, almost two years later, the ASRG and MARID working groups are still not significantly further, have not yet defined a standard, and the MARID proposal was not significantly better than RMX. On the contrary: RMX was designed in 2002 to also protect against wrong bounce messages, which are more of a problem than spam itself for many domains. The current MARID proposal SenderID (made by Pobox and Microsoft) did not protect against those bounce mails anymore, without good reason. And it was still months away from being finished and tested. It was not even clear what is to be authenticated.

The inability of the IETF to produce a standard and to protect SMTP did not only cause a delay of more than a year. It gave Microsoft the time to apply for patents on the method. Microsoft did not invent that mechanisms and the act of applying for that patent is to be seen as a malicious act of hijacking the technology and preventing any form of free and open source implementation. This was one of the main reasons for MARID to fail. Due to their commercial orientation the group chairs focussed on the Microsoft/SPF/SenderID proposal only and suppressed any other option. When Microsoft raised their patent claims, many authors of open source email software and Linux distributors declared to be unable and unwilling to support that method because of the abusive license attitudes of Microsoft. On the other hand Microsoft declared that they'd rather let the whole technology die than to resign from their licensing demands.

Although RMX was clearly not only prior art but also well known to Microsoft, I as a private person cannot afford to defend against a company like Microsoft.

It is currently very difficult to form a standard in this situation. It would take the US government's intervention to stop the abusive attitude of Microsoft. And it is an open question who could form such a standard after IETF failed to do so.

It will however require to restart from scratch and thus again take months.

Commercial hindrances

Commercial software (MTAs, MUAs, Mailing List Processors, Forwarders, DNS) turned out to be another major problem.

Unfortunately the SMTP standards are not that precise and do not define exactly, how MUAs, listprocessors etc. have to behave. Many commercial email and DNS products suffer from poor implementations, proprietary protocols, and arbitrary user interfaces. Some of them use proprietary protocols internally which are not really compatible with SMTP. For example, the SMTP envelope sender address is one of the most important thing in mail autentication. Most commercial MUAs can't be configured to display that address, and some commercial MTAs simply ommit this information. The sheer variety of arbitrary and proprietary interpretations of email services makes it difficult to fight spam and is one of the main reasons why IETF failed. The influence of commercial parties trying to make a protocol compatible with the flaws in their products and services defeated success.

It is a shame that commercial interests and workarounds for poorly designed and implemented commercial software are more important than security.

Q1: Will authentication result in decrease of spam

See my first two comment above (page 2). Authentication technology will not reduce spam. It will reduce the amount of forged messages (after all, that's what the term authentication means).

Since spammers currently do forge, it will temporarily cause a certain decrease, but this won't last. As I pointed out, it will have to be used in conjunction with some other technology, which is not yet available and won't be applicable on a world wide scale.

Q2: Modification of Internet Protocols

None of the currently existing proposals requires any change in layer 3 (IP) or layer 4 (TCP/UDP) protocols, although modifications on layer 3 (IP protocol) might allow improvements of security.

Currently two different ways of storing records in DNS are under discussion. Some propose to store the authorization records in DNS TXT records, which is bad design but can be used with DNS servers currently running.

The technically better approach is to use a new DNS entry type. This is not a modification, but an extension. The DNS protocol is designed and supposed to be extended by new entry types. Under normal circumstances this would require just an update of those DNS servers acting as domain primaries. But flaws in Microsoft DNS software make it difficult to do this.

Another problem is the shortage of IPv4 address space wherefore NAT is widely used. A whole company network hidden behind a single NAT address does not allow fine granularity of authorization. IPv6 should be enforced.

Q3: Software and Hardware required

RMX-like authentication methods do not require new hardware (except maybe for DNS servers with more main memory). Depending on implementation details it would require an update of the DNS servers and new MTAs to perform the check.

Methods based on cryptography are different: In contrast to normal communications, the attacker (spammer) does not need to compromise a particular key. Worms, Viruses and Trojan Horses could collect the secret keys from infected computers. Even if you assume that only about 1% of all computers are infected, there would still be tens of thousands stolen secret keys available to spammers, allowing them to forge emails.

From my point of view this is the main reason why cryptographical methods are not appropriate to protect against spam. To prevent this a cryptographic hardware device would be required. This is unfeasible because too expensive and incompatible with some legal systems.

Q7: Requirement of open standards

Any proposal must be an open standard. The world wide email system is mainly built on open source software. The authors of open source software would not adopt an non-open standard. They already refused to accept SenderID because of Microsoft's patent claims.

Furthermore too many countries would not accept a non-open standard dictated by the Unites States. An open standard is the only chance to have a widely accepted standard.

Q8,9: Whether any standard is patented

As explained above, Microsoft has applied for patents on the technology of RMX, thus covering virtually all of these proposals. I do consider this as an act of hijacking and robbery of intellectual property.

However, virtually all proposals must be considered as being patented because of this situation.

Q10: Forwarding Services

There is no concise answer to this question because of the variety of those services and for reasons of security.

Basically, any forwarding service is required to take over responsibility of the message transfer, otherwise any spammer could pretend to be just a forwarder. RMX introduced the concept of trusted forwarding.

Q11: Impact on mobile users

Regular mobile users will not be affected since they usually do not do SMTP MX-delivery. They usually deliver their outgoing email to a fixed relay address after authentication. This mechanism is not affected by anti-spam mechanisms. For users who still want to deliver directly from mobile computers (or dynamically assigned IP addresses), RMX supports DynDNS entries.

There is nevertheless an asymmetry in mail delivery that should be fixed to alleviate spam protection. Most mobile users do download email through POP, but upload with SMTP, which requires firewalls etc. to allow SMTP. A better way would be to introduce bidirectional POP and to completely eliminate SMTP for mobile users (and dialin users with dynamically assigned IP addresses).

Q12: Impact on roving users

This is not a trivial problem, because if roving users could use third party submission services, the spammer could also forge and pretend to be a third party for a roving user.

A better approach would be to eliminate the need to use third party submission services (e.g. by bidirectional POP).

Q13: Impact on mailing lists

As far as I can tell, regular mailing lists will not be affected and can proceed as usual, because the mailing list processor itself is the sender of the distributed messages and can easily be authenticated.

Unfortunately, there are some broken mailing list processors that will cause problems and need to be replaced or updated.

Q14: Impact on outsourced email services

Current proposals do support delegation.

Q16: Impact on web-generated mail

Web-generated mail is generally just a special form of outsourced email services. The domain owner needs to authorize the web service provider (or delegate authorization).

There are special forms of web-generated mail where a foreign web-server without relation to the email writer is generating the message, e.g. greeting card senders or customer care web forms. These web server usually ask for the writer's email address and use it as the sender address of the message. These services will not be possible anymore without change, otherwise the spammer could forge by claiming to be a web service sending on behalf of someone else.

Those services would have to send under their own domain and responsibility.

Q19: Impact on ISPs not participating

Mail authentication does not necessarily involve the ISP. The ISP is only involved if it provides DNS and SMTP services. If so, the ISP would have to participate. Otherwise, any spammer could pretend to be an ISP not wanting to participate. The ISP could then face incompatibilities and inadequate license fees or terms. The authentication mechanism might be conflicting with the ISP's infrastructure, e.g. if the mechanism is under patent of Microsoft and the ISP uses open source software as described above.

That's another reason for keeping the "authentication regime" open and patent free.

Q20: Adoption in reasonable amount of time

Adoption is not possible in reasonable amount of time anymore, since the IETF has already wasted more time than what would have been reasonable.

An authentication system can be widely deployed only if it does not make use of cryptography, because too many legal systems do not accept cryptography and too many email users do not have the infrastructure required. Too many email users will never accept a system where they do need an license from Microsoft, as we have recently seen and what caused the IETF to fail. Any other attempt will fail the same way.

For fast adoption the system must be open, with public specifications, simple, robust, free, independent from american companies and government, compatible with any legal system and should not make use of cryptography.

Q21: Delay of mail transport

Every known method will more or less delay the transport of email. It will at least require an additional DNS fetch, and maybe further communication (e.g. RMX++) or additional computations (methods based on cryptography). Keep in mind that email is not an interactive service like web or chat, it is a store and forward service and thus may take some delay by definition. Time delay will be within seconds or milliseconds and thus below human perception.

Spam is often more than 50-70% of the ISP's or user's email traffic. I expect the time delay of authentication and authorization to be at least or even more than compensated by the reduction of the amount of spam and thus traffic.

Q22: Impact on anonymous political speech

Actually, spam protection will have impact on anonymous policital speech by email. But anonymous political speech does not include the right to force everyone else to listen and to abuse someone else's infrastructure. Spam protection does not necessarily mean to reject anonymous email. It is up to the recipients preferences to sort it into a specific folder and to check it once a week for interesting contents.

So the main impact is that it turns the decision of the speaker whether to speak into a decision of the recipient whether to listen.

However, I do not consider anonymous political speech as a valid argument, since the email system does not support anonymous speech anyway. You should not confuse the possibilty to exploit a flaw in SMTP to forge email with 'anonymous political speech'. From the security engineer's point of view, exploiting the lack auf authentication by forging sender addresses is far from beeing anonymous speech. It is just an attack.

If you do consider anymous political email as desirable, then you need to reengineer the email system and to state this as an explicite property.

You still might encourage anonymous remailers which could prevent forging and detect and block spamming, while still allowing sender anonymization.

Q23: Antitrust laws

As I pointed out, there is already a severe problem with Microsoft claiming patents on other people's inventions. The scandal about the IETF MARID working group might have been caused by some kind of cartel or arrangements in the background.

It is absolutely necessary to protect the development of an anti-spam system against that kind of influence. The US government is the only party strong enough to do that.

Q24: Can a hacker compromise?

Yes, of course (but RMX++ provides counter measures).

The current state of the art does not allow to design a mechanism completely unbreakable and suitable for world wide deployment. The reason is simple: There are too many computers with vulnerable operating systems and applications. If the user sitting at the keyboard can send an email, then any hacker taking over control of that machine can do it as well. If you want to make the email system completely hackerproof, then you need to exclude all Windows machines from the email system in a first step.

An exception is RMX++, which allows a central machine of the domain owner to be queried for every single email. It can detect if a business machine sends more emails than a given threshold, sends after business hours, or sends to unusual recipients and automatically remove the authorization for that machine, which would effectively block spam robots and hackers.

Another reason is the vulnerability of the Domain Name System DNS. A hacker could poison the zone tables. Cryptographical certificates issued by a central CA (similar to SSL certificates) would avoid that, but the costs and the overhead would be far to high. Again, although the flaws of DNS are known for years, IETF failed to provide a practical replacement.

I want to point out that cryptography-less systems are more robust in my opinion. Since it is too expensive to equip every sending MTA with a crypto device, the secret keys would have to be stored in the filesystem. Plenty of viruses, worms, and trojans exist which systematically gather secret keys. There are several millions of domains. Even if only one percent of the secret keys of those domains were compromised - which is a rather low estimation - this would still mean several tens of thousands of compromised domains.

The problem is that spam does not fit into the classic security model of authentication, integrity, and confidentiality. The attacker just needs to compromise any key of any human on the world - a user group the attacker himself belongs to and that will always have plenty of keys compromised.

Another problem is that any DNS based spam protection system is vulnerable to denial of service attacks. If an attacker manages to modify a domain's DNS entries, a domain (e.g. a large company) can become unable to send any email. On the other hand, this not really a new vulnerability, because blocking MX records would have a similar or even worse impact.

Q25: Will it prevent "phishing"?

This question is not that easy to answer, since phishing and identity theft are not preciseley defined.

A major problem is that phishing is not limited to email. Any Email protection will not help against web based phishing. A second problem is that even email based phishing methods are not always based on forgery.

However, I do believe that my proposal about blocking emails from gTLDs such as `.com` in general (page 2) would provide significant protection against phishing. Phishing is often initiated by email from domains that just look like a third party's domain (bank etc.). In most cases phishing is undertaken from other countries.

Blocking email from gTLDs such as `.com` and requiring country specific domains with authentication would make it impossible to hide the country the attacker is attacking from. People should usually get suspicious if they receive an email from their bank but from a different country. A citizen of the United States could always expect email about private affairs to come from within the `.us` top level domain, which is a simple and easy rule. Even unexperienced and simple minded people can be able to cope with that. This is impossible with the `.com` domain. It is then up to the US legislation to ensure that every `.us` domain can be tracked back to a responsible person.

Q27: Cross-border implications

Methods based on cryptography may not be allowed under all legislations. Methods based on reputation and accreditation might not be acceptable under some legislations or not meet the mentality of other nations.

My proposal to limit mail sender addresses to country code Top Level Domains (page 2) help avoiding cross-border implications by leaving it to every country how to deal with spam according to the country specific legal system.

Q28: Need for internationally adopted standards

See my comment to question 27 above and my proposal (page 2).

Q29: How can the summit support efforts

After the IETF MARID working group was closed at Sep 22, there is currently no working group to define a standard. Failure of IETF to develop a protocol dropped everything into a vacuum.

By far the most important support the summit can give would be to found or establish a new working group.