Overcoming the New Paradoxes of Spam

Kenneth Neil Cukier
Technology Correspondent, The Economist

Anti-Spam Session
4th ASEM eCommerce Conference
22 February 2005; London, UK

Good Afternoon.

Let me first begin by thanking Martin Boyle, Jean-Jacques Sahel and everyone else at the UK Department of Trade and Industry for their work in putting together such an informative two days. Often spam issues seem intractable, but after the excellent contributions from all the speakers today, I feel we are in better shape to tackle many of the challenges with which we're faced.

Before I introduce the panel and begin the discussion, I've been asked to make a few remarks. I'd like to take a moment to take stock of, and indeed respond to, a few of the most interesting things I've heard today from other speakers -- that this panel can use as the basis for our broader discussion, on where we go from here with spam.

I've structured my thoughts around three observations, which may be contrary to the customary views on the topic:

* First -- We will always have spam, and that this is actually a good thing.
* Second -- The solutions we adopt risks hurting the very thing we're trying to protect.
* Third -- The chief obstacle to overcome spam isn't technical or legal, but mental: we need new ways to think about spam, in order to limit its harmful effects.

I.

So first: why we have spam, why we always will, and why this is actually a good thing. Spam exploits two fundamental vulnerabilities of the Internet: the ability of users to hide their identities, and a change the traditional economics of communications from the cost being borne by the sender, to being shared by the receiver of the communications.

The reasons for this goes back to the origins of the Internet, specifically, its precursor, the ArpaNet. It was designed to enable data communications that let different systems talk together, without a single point of failure. To do this, it had to be more efficient than the classic telephone network -- and the way to do that, was to get rid of maintaining intelligence at the center of the network, and instead place it at the network's edges. In short: decentralize it. At the time, the physical infrastructure was secure, so the protocol didn't need to be. Think about it: the hardware of the Net was protected -- it was the US Army, after all -- so the software didn't have to be quite so secure.

Keep in mind that this is a time when there are only a few hundred people online, they are almost all American researchers at academic institutions and government research labs, and everyone need permission from the Pentagon to get on the network. No one expected the network built in 1969, and then upgraded in 1977 with the advent of TCP/IP, to be almost the same that exists today. Indeed, in 1996 *The Economist* called the Internet "The Accidental Superhighway." Some people raised security concerns. David Reed, an early Internet pioneer, then a graduate student at MIT in the early 1970s, was prepared to write strong security into the protocols, but was told by the engineering community that it just wasn't a priority. And so the initiative was abandoned.

This history is important because it leads to two points: First, why the Internet succeeded against its rivals. The Internet Protocol wasn't the first or best or most-funded data-networking system. In fact, for most of its life it was the underdog, facing better rivals, backed by the world's biggest companies. Names like France's Minitel system (on the X.25 standard) or IBM's Simple Network Architecture, AT&T's Token Ring, Digital's DECnet, and many others litter the landscape of lost technology.

The reason the Internet succeeded is because of its fundamentally different approach to networking: an "open" versus "closed" network design. Control was placed at the end points, with the users, not in the center, with the network operator. For France Telecom, the closed system meant that every single piece of Minitel traffic could be monitored, measured, metered, and billed for -- and it was, *très cher*. It meant that new features and upgrades had to be decided and implemented by the owner of the network, not its users. And innovation was limited to what a few hundred data engineers in Paris and Grenoble came up with, under the dictates of their corporate masters.

The open approach is the opposite. Users can decide how to employ the network, identify new uses, and innovate. That's how a hot new service like instant messaging can be created in the 1990s and spread like wildfire, or why voice-over-IP services can be established without the approval of the telecom operators -- who would otherwise have an incentive to block it since it jeopardizes their revenue. Or how Napster can emerge literally overnight, and change the music industry.

The second reason why the history of the Internet and the differences between open and closed networks is important, is because of the vulnerabilities that this openness creates. The lack of centralized control that leads to an ability of users to mask their identities. The decentralization also diffuses the cost of communications, where senders pay a bit -- but so do receivers, and thus things like spam can proliferate, since it is the receivers that bear much of the burden. These two features have led many people to call for the Net to be changed in order to remedy what they perceive as shortcomings.

II.

This brings me to the second point: in trying to eradicate spam, we risk undermining the very medium we are trying to save. Many technologies, in trying to plug the holes of the

Internet, actually do much more: they effectively "close" the openness of the Internet, which as I've tried to suggest is the reason for its very success.

For example, consider filters. They are always overbroad, and by their nature block out legitimate communications. Indeed, they have to be this way. To be even remotely effective, they must permit false-positives in order to prevent false-negatives; in other words, the systems are designed towards overkill: they would rather have some genuine email get deleted than let some spam get through. As a result, email as a medium is less reliable -- in this case, not because of the "bad guys" but because of the "good guys."

The same problem holds true with the notion of "black-lists." They are overbroad, never truly up to date, and are inherently designed to block out too much rather than not enough (lest they be accused of not working at all…). Not to put too fine a point on it, but to offer one example: the International Telecommunication Union -- the United Nations body that manages global communications -- for the longest time didn't accept email from my personal domain, cukier.com; the messages would bounce back as unsendable, so I wasn't able to communicate with officials there. Clearly, this is unacceptable.

As for the opposite of black-lists -- "white-lists," where only email from approved people can be received -- they inherently prevent communications from friends-of-friends, or the "serendipitous stranger," among the most common types of communications. One approach, used by one of the world's foremost security experts Ron Rivest of MIT, is to require unknown emailers to read a short note and click a link that send a message to another email address as a way to ensure that it is a real person emailing. If this is done, then the email can be received. In theory it is fine; in practice, if universalized, it means that we would need to expend twice the effort each time we want to communicate with someone we don't know -- which is totally impractical.

In short: in trying to navigate the technical "solutions," we spend more time than we otherwise need to, making email a less efficient medium.

But even worse than these approaches is an effort to add more "centralization" to the Internet as a way to remedy its ills, from spam to phishing attacks to porn and press censorship. This is a step-back in time -- like returning to the telegraph to overcome heavy-breathers on a telephone line. A matter of control. This view of re-casting the Internet in the image of the telephone system is actually happening at the ITU, under the framework of the "Next-Generation Network" (or, NGN) discussions. There are many parts that are extremely valuable, but there are other aspects that are very misguided.

I raise this because it is important to reiterate that the Internet was successful because it was decentralized and open. It meant that costs of access were low; almost nothing, and led to easy, ubiquitous access. (When the Net first arrived as a mainstream medium, people used to call it "free" because it was so inexpensive -- emails weren't measured for kilobyte size, or counted in number, or metered in distances-sent, like phone calls…).

We take that for granted today, but it was novel in the mid 1990s. Companies that tried to apply classic telco charging models onto the medium were beaten back. I personally recall having to indignantly pay my network provider CompuServe ten cents to open any email that came from someone other than a fellow CompuServe holder in the early 1990s. After much outrage, the company backed down. (This was an era when I received around 5 emails a day… imagine what it would have meant for the evolution of the Internet if the practice was universalized.) The open approach won -- and the result is that the medium won, too -- it became more widely adopted than it would have if it embodied the closed-approach to network design.

This technical openness means that no one can "control" the Net commercially because it allows for a diversity of business models. This diversity is like capitalism on the economic sphere or democracy in the political sphere. It is a more efficient method of organizing resources, and through trial and error usually the best have chances to emerge. This of course is also the fundamental precept of scientific experimentation, and what has led to the Internet's tremendous force for innovation not only for the network itself, but for all aspects of the world that incorporates the Internet into its operations, from retailing to decoding the human genome. (And it is not an accident that the Internet would embody this enlightened approach: it was designed by academics, after all!)

We may complain about the openness of the network and the problems this creates, spam first among them, but when you think about it, they devolve into the same arguments that faced all civilizations that confronted the open society and its enemies. Free speech is problematic. But rationalists know that it far better to deal with the problems in very surgically precise ways where a clean cut is possible, than start from the position that speech ought be restricted. Likewise, the open network has drawbacks -- but the benefits it brings ultimately outweigh them. That does not mean we should turn a blind eye to the problem, but it weighs on us to tailor solutions that don't harm the medium itself.

If we were to close the Internet in order to keep it open, it would be akin to cutting off a nose to spite one's face, or in a more tragic formulation, to destroy the village to save it. Shall we cut the Net to spite the spam?

III.

This leads me to my third observation, that the chief obstacle to overcome isn't technical or legislative, but mental: we need new ways to think about spam, in order to limit its harmful effects.

Churchill once said -- and I paraphrase him badly -- 'If a problem doesn't have a reasonable and clear solution, it may be the case that it isn't a problem to be dealt with, but a reality that must be lived with.' Thus: spam. The problem today is *not* that spam is a nuisance. It is. But the problem with spam is that, firstly, it is a symbol of the vulnerabilities of the network generally. And secondly, it is a carrier for more menacing dangers, like viruses or phishing exploits.

Thinking about spam in this way helps us consider possible ways to tackle the problems of spam -- that is, not tackle spam, but the problems associated with spam. As a first step, we need to separate the symbolic vulnerabilities and treat each issue independently. As veteran policymakers know, trying to achieve more than one policy goal with a single instrument usually leads to failure, just as the fellow that chases two rabbits ends up with none. So targeted approaches, not sweeping ones, are called for. Second, much of the unease about spam is really a concern about the illegal activity for which spam is the vehicle. If we crack down on that -- whether its fake drugs or identity theft -- much of our anxiety will be assuaged.

This is not a call for passivity. But it is not a call to action, either. It is a plea for the appropriate action. We need to remedy the harms of spam through legislation, litigation and technology. But we need to understand that we have to live with it to a certain extent. And that this is a worthwhile burden of having an open, not closed, network -- in the same way as we have an open, not closed, society. It begets problems, but also freedom.

* * *

I'm convinced that in ten years' time we will still have spam. If there were an easy answer to it, clearly we'd have it by now. That said, I'd suspect that in a decade we won't still be talking about spam. Instead, we'll be discussing new problems the network engenders, and we'll be living with -- and taking for granted -- all the benefits the Internet's openness brings, as well.

Thank you.