PoET

The Cambridge-MIT Institute

oiioiioi
oiioiioi
oiioiioi

Future Internet Design Workshop

27th and 28th April 2006

Oxford Internet Institute, 1 St Giles, Oxford

Thursday 27 April:

10:00	Welcome, background, goal of meeting.
10:30	Case study 1: Trust in content: authorship and authenticity (DDC)
11:30	Case study 2: Trust in email (JZ)
12: 30	Lunch
2:00	Case study 3: Identity in the net-necessary or harmful? Talk: David Clark, 30 minutes Discussion 30 minutes
3:00	Port numbers-an analysis and discussion Talk: Betsy Masiello
4:00	Case study 4: Managing location information (JZ)
5:15	Wrap up first day. How are we doing, assignments for tomorrow
5:30	Adjourn
7:00	Dinner, Forte Room, Pembroke College

Friday 28 April:

9:15	Thoughts after a night to think: revise agenda, proposals for the day.
9.30	Design exercise 1: Building an instant message system with assured logging. Talk: David Clark: 30 minutes Discussion: 30 minutes
10.45	Higher level discussion on IM continued: What if we did this? Second order effects?
11:15	Discussion: peer-to-peer judgments and reputation systems
12.45	Lunch
1.45	What did we learn about design? What did we learn about collaboration? What did we learn about process?
2.45	Wrap-up
3:15	Adjourn

Session Outlines

The goal of this workshop is to explore a new space of multi-disciplinary design, and the related question of process—how can a multi-disciplinary team function effectively to contemplate design alternatives for a future Internet?

The workshop will consider several case studies, which will be the basis for a series of mini-design sessions. Each case study will describe a set of design alternatives, which we will consider in terms of both the technical options and the social or economic implications.

Case study 1: Trust in content: authorship and authenticity

Speaker: Dr David Clark, Senior Research Scientist, MIT Laboratory for Computer Science

Today, we establish the authorship (or authority) of a piece of information indirectly. The way we confirm that a web page is (for example) created by CNN is that we download it from CNN. The confirmation of the authorship is associated with how we got it, not with the object itself. If A downloads a web page and then send it to B, there nothing about the web page itself that confirms to B that it is legitimate. Should the Internet include a way for authors to sign Web pages and other net documents, and how should this namespace be controlled?

There are lots of reasons why it would be important to validate the authorship of a piece of information independently of how one gets it. Today, bits of information are sent around from person to person as email attachments. The research community has proposed a number of creative schemes, such as peer-to-peer retrieval system, to support efficient and robust information dissemination. And as we more and more depend on search engines such as Google to find things, we have less and less confidence about the origin of anything we download. But today, none of these schemes can confirm the validity of the information they propagate. To mitigate these problems, what Is needed is a scheme that associates the authority with the object, not how it is fetched.

It is not difficult to conceive a scheme to sign documents with some sort of encryption scheme. The hard part of this problem is to devise the namespace in which the authors are named. Such a system has to resist forgery, be easy to manage, and easy for users to understand. Today we use DNS names and public key certificates to verify the source of data, and at first blush, we could just use those same names as names of authors. However, the ongoing debate about governance of the DNS suggests that we might consider designing a name space that sidesteps some of the tough political tangles that have snared the DNS.

One choice for naming is that there is some sort of registry of author names, similar to the registry of domain names. Provided the registry is trustworthy, this approach gives the receiver of an object some confidence about a page from a previously unknown author, but raises all the issues we have today with control and governance of the registry. Another approach is that names are not registered at all, but are just validated pairwise between sender and receiver. That is, a receiver comes to trust a sender by developing a relationship with that sender. The names provide continuity, but not the trust in the relationship. This concept is similar to the discussion about “bottom-up” signing of email.

Case study 2: Trust in email

Speaker: Professor Jonathan Zittrain, Professor of Internet Governance and Regulation & Director of Graduate Studies, Oxford Internet Institute

In the Future Internet, when person A sends an email to person B, what level of information should B expect to have about A? With what level of certainty? How would acting to strengthen confidence/knowledge of identity in the Future Internet affect users? Would the benefits of stronger identity outweigh the possible negative consequences?

Today, users of email identify themselves to each other using their email addresses, strings such as user@example.net. Between cooperating users, these work well, but they can be forged. So a certain amount of spam is sent using a forged sender email address, and the mail classified as "phishing" has false sender email addresses. There have been calls to "improve" this situation, and give person B a higher level of confidence that person A is actually person A, and a higher level of knowledge as to who person A actually is.

If it is generally agreed that identity needs to be strengthened, further questions remain as to how this would be accomplished. One option is that all users should be issued a "top-down" identity—that is, an identifier issued by some trusted party, who is in turn validated by some higher-level trusted party, perhaps leading to a chain of trustworthy identities that can be traced back to a nation-state. This identity could be used to identify the sender to the receiver. In a system like this, the question is: What attributes of identity are included (and verified) by this issued identity? Would it be just the given name of the person? Name, address and years at that address? Age? Gender?

Another view is that we use "bottom-up" identities. In a system like this, users issue themselves identifiers. If properly designed, these cannot be forged, but of course since they are self-issued, they do not tell the receiver anything about the sender. All they allow is a sender to prove that he is the same sender as in the last message. They allow a user to connect a series of interactions, and "get to know" someone by that means.

Technical form of the questions: Should we move toward a regime of mail signed with public/private key pairs? If so, should we presume a single PKI, several unrelated PKIs, or self-signed certificates. How can we present identity information in the user interface of the mail system so that the user can make sensible decisions based on the information?

Case study 3: Identity in the net-necessary or harmful?

Speaker: Dr David Clark, Senior Research Scientist, MIT Laboratory for Computer Science

In the future, to what extent should an observer "in the network" be able to identify the sender of a packet? What are the social and economic implications of different degrees of identity and anonymity in the network? How would the presence of identity information change current uses of and interactions on the Internet? Should this ability be used to develop a policing system or "surveillance cameras" for cyberspace?

Today, the source IP address in the packet gives some hint as to the identity of the sender. Precisely, it indicates only the network location (not the physical location) of the sender; however, this information can often be translated into some better signal of identity. This link to identity may get much weaker. With increasing mobility of end-nodes, the IP address will become more of a transient indicator of location. The location may change rapidly over time, even for an ongoing connection. Given this trend, there is essentially nothing visible in a packet that can link it back to the identity of the sender.

There will, of course, be identifying information handed back and forth between the end-points of a connection. At a minimum, there will need to be some sort of identifier that the end points use to keep track of each other as their location changes. Some end-points may demand very robust identification from the other end-points. But while this information is carried among the end-points inside packets, it may not be visible—it may not be in a standard recognized format and it may be encrypted. Further, it may not be in all the packets of the connection. So while the end-nodes may know each other's identity, observers in the network may not. Given this fact, a "surveillance camera" in cyberspace may be able to see nothing except encrypted packets going between transient locations. We could choose to include explicit identity information in packets, but how might we

minimize the chilling effect that is associated with loss of anonymity and analogs such as identity cards and travel papers?

Hence the question: are there significant reasons for packets to carry some sort of identity that is visible "in the network"?

Technical form of the question: Should there be a field in the packet header that identifies the sender of a packet? If ISPs want to give different users different services, what information in the packet should be used to signal that discrimination—session setup or packet state? If the packet had an identity field in the header, do we need to fully specify its semantics, or would it make sense to use the field in different ways at different times and places?

Port numbers-an analysis and discussion

Speaker: Betsy Masiello, MSc Student, Said Business School, University of Oxford

Today, the Internet names services (such as Web or email) using "well-known ports"—numerical indices that are statically assigned to each application and service at design time. Since these port numbers are included in each packet, this permits any observer in the network to determine what application is being used. And since these numbers are statically assigned, an attacker can easily launch an attack against an application on a given host. It need not be this way--an alternative would be to design a new mechanism for "service rendezvous", and to use random port numbers to identify connections. What would the implications be if packets did not reveal what applications they were associated with, and security tools could not associate applications with ports.

There are implications for economics and security if it is not possible to tell what application the users are running by looking at the packets "in the network"? ISPs observe port numbers to build models of what their users are doing. This information may help with capacity planning, and is the basis of discrimination among different classes of users that wish to use different services.

From a security perspective, well-known ports are a fundamental component of both attack and defense. Attackers combine host addresses and well-know ports to explore whether a host is running a version of the service with a known vulnerability. Firewalls respond by blocking access to specified port numbers. If port numbers were random, this could essentially eliminate the value of the attack known as port-scanning, but it would change the whole security landscape by limiting what firewalls can do based on packet inspection.

So the question: should a future Internet employ a different scheme for session rendezvous and connection identification.

Technical form of the question: Should the Internet move to a mechanism such as DNS-SERV to assign port numbers to services on hosts? Should the port field be increased, perhaps to 32 bits? Should we move to an application-specific mechanism for rendezvous and connection management? Should the port number be encrypted in some way?

Case study 4: Managing location information (JZ)

Speaker: Professor Jonathan Zittrain, Professor of Internet Governance and Regulation & Director of Graduate Studies, Oxford Internet Institute

There is no notion of physical location in the design of today's Internet. But with the increasing use of mobile and wireless devices, physical location will become of increasing importance. Today, the cellular system has been required to determine the location of every cell phone so they can respond properly to a 911 emergency call. Physical location is integral to some emerging on-line socialization,

flirting and game-playing. It seems inevitable that physical location will be central to many future Internet applications. So now is the time to consider how a system for managing physical location should be designed.

How is location derived? Today, in the Internet, addresses are used as the starting point. Addresses are given out to Internet Service Providers, who give them in turn to customers. Most ISPs operate within a physical locale (which may be large in the case of a national ISP), and most ISPs know where their physical circuits go. If they know that a customer is at the end of a physical circuit, they can make a reasonable guess as to where that customer is. This sort of inference is used to prevent auction services from displaying Nazi memorabilia to citizens of France, for example. Services such as Google give different answers depending on what country you are in, and this requires some sort of guess as to physical location based on IP address.

In the future, there will be many more ways to determine location. More and more end-node devices have GPS receivers, so if they are outside (in view of the open sky and the GPS satellites) they can compute their own location. The question then is whether (and to what parties) they will pass on that information. GPS information is under the control of the end-node, not the ISP. But the wireless service providers are also getting better at locating their end nodes. Wireless systems can attempt to geo-locate end-nodes using the radio signals, and triangulating.

The question, then, is what will operators do with this information? One obvious answer is to sell it. This may trigger calls for regulation of the use of this information. This will in turn raise question about what third-party players who attempt to guess a location will be allowed to do with their information. To help visualize what the issues might be, here are some possible scenarios of usage:

Case 1: Use of location in an ongoing interaction. In this situation, some number of parties are already in communication, and one of them attempts to determine the location of another. Since they are in communication, each end knows something about the other—at a minimum an IP address. There may be other information available. This is the situation with country-specific modification of services like Google or E-bay. In these cases, the question about location is asked about a specific node that is already known.

Case 2: Use of location to find possible candidates for interaction. In this case, the interaction starts with a request to determine a set of nodes that are within some locale, after which an attempt is made to communicate with them. Examples of this include sending a message to anyone on a beach that a tsunami is coming, or sending an advertisement to anyone nearby about a sale going on. The social value may vary widely, and there may be calls to allow this sort of location-based search only under certain circumstances, which will have to be debated.