Security research challenges for future mobile systems

Presentation for **Research Network for Secure Australia** *meeting: QUT, Brisbane, 7th July 2005*

Chris Mitchell

ISG, Royal Holloway, University of London

http://www.isg.rhul.ac.uk/~cjm

Introduction

- Security and privacy research topics of importance to future mobile systems are reviewed.
- This is inevitably a very personal and Europe-centred view (apologies for obvious omissions).

Research themes

- There are two major themes running through this presentation:
  - *Mobile system security/privacy*. There are many well-known problems here, particularly in ubiquitous / pervasive / ad hoc computing scenarios.
  - *Applications of trusted computing*: It is possible (likely?) that the existence of trusted functionality on future mobile platforms will help to solve many apparently intractable security and privacy problems in the mobile world.

1. Mobile systems privacy

- Many of the (mobile) devices providing the future ubiquitous environment will be personal devices; they may thus 'leak' personal information to each other.
- In a context where devices cannot be assumed to belong to a single trusted domain, there are thus major privacy issues.

Interrogation of mobile devices

- Communications protocols for mobile devices inevitably require some form of routine 'polling'.
- Responses to polls (e.g. from a network access point) need to contains some kind of identifier, e.g. a network address.
- Thus can be used to 'track' devices, and potentially track the location of device owners.
- Possible solutions? GSM/3GPP use temporary identifiers (pseudonyms) distributed in a way that prevents linking. Provide confidentiality protection for exchange.

Use and abuse of authentication

- Authentication of a device can pose a denial of service threat.
- For example, if protocol requires one device to store state and/or do computations, repeated fake requests can cause memory/processing exhaustion.
- Possible solutions? Use stateless protocols. Require requester to do at least as much work as the responder.

Location information use/privacy

- Service providers in a ubiquitous computing environments may wish to provide services based on user location, e.g. targeted advertising, emergency services, broadcast blackout, …
- Owner of computing device may wish to restrict dissemination of such location information.
- How should this be controlled?
- Possible solutions? Anonymity. Mandatory inclusion of policy data with location information. TTPs.

Denial of Service versus privacy

- In any protocol it seems that one party has to reveal their identity first. This argues that (mostly) the requester of service should reveal their ID last. (P2P an exception?)
- However, this potentially increases the risk of Denial of Service attacks against the responder.
- Indeed, more generally, the tension between DoS-resistance and user privacy has been noted by a number of authors.
- Possible solutions? New ideas needed?

Accountability versus privacy

- Anonymity to protect privacy may cause major problems in making users accountable for their actions.
- An audit trail (present to provide accountability) is useless if the real owner of a pseudonym cannot be determined.
- Possible solutions? New ideas needed?

2. Mobile systems security

- There are a huge range of new challenging issues which arise from the use of mobile devices making up part of a pervasive computing environment.
- We touch on just a few…

Ad hoc working relationships

- Major issues exist in establishing ad hoc working relationships between mobile devices (in absence of pre-existing security infrastructure).
- Initial trust setting is a major issue.
- Desire for zero or near zero configuration overhead – automatic security initialisation is required.

Automatic address assignment

- In an ad hoc network, newly admitted devices will typically need to be assigned a network address (or addresses).
- In the absence of a fixed infrastructure this is problematic.
- Solutions can easily lead to the possibility of denial of service attacks.
- Just one part of 'zero-config' problem.

Routing in ad hoc networks

- Many protocols designed to enable distributed routing in ad hoc environment.
- However, such schemes are prone to a variety of attacks, including ‘selfish’ behaviour.
- Protocols are needed which can address a variety of possible security threats without requiring a global security infrastructure.

Location of security functionality

- Where security functionality is located in a protocol stack can have a significant effect on security provision, including:
  - Encrypting at the application layer will not hide any of the lower layer addresses and routing information. May also cause problems for firewalls.
  - Integrity protection at individual link level will not provide end-to-end integrity protection.

End-to-end versus point-to-point

- Need for security between service provider and service consumer argues for end-to-end authentication.
- Need for control of access to resources, e.g. network access, argues for point-to-point authentication.
- If both provided in 'unlinked' way then man-in-the-middle attacks can become possible.
- Great care needed in combining protocols at different levels in protocol hierarchy.

Protocol statefulness

- As mentioned previously, protocol state can be used as a means of launching DoS attacks.
- 'Accepted wisdom' is to require protocols to be stateless, at least for responder.
- However, there is an efficiency cost (state must be shipped in protocol messages). It also either requires synchronised clocks or regular key changes (a bit like state).

Key management and security infrastructure issues

- Use of crypto requires either shared secret keys (using symmetric crypto) or trusted copies of public keys (using asymmetric crypto).
- Shared secrets can be set up via a mutually trusted TTP.
- Public keys can be obtained via public key certificates as part of a PKI, although trusted means to verify certificates (CA public keys) required.

Heterogeneous networks

- The pair of devices may not share an online TTP (or even share 'trust-connected' TTPs).
- Public key crypto (and PKI) looks more promising, but it is still necessary to have mutually verifiable certificates. Finding certification paths (where one CA certifies the public key of another CA) could be infeasibly complex for a bandwidth-limited device.
- Possible solutions? Delegated Path Discovery/Delegated Path Validation.

PKI interoperability

- Finding a certification path is by no means only problem with using PKI.
- Certificates issued by different CAs (with different policies) may 'mean' different things – e.g. different liability protection, different ID checking for certificate issue, etc.
- Certificate status management systems may vary.

ID-based cryptography

- One possible solution to key management problems is used of ID-based crypto.
- Here a user public key is derivable from a user identifier (possibly plus other data).
- Requires TTP to issue private keys (and TTP public parameters to derive public key from ID).
- Hence we have major interoperation issues if two devices served by different TTPs.

Mobile code security

- Mobile code, i.e. executable code which is downloaded from one device to another (e.g. agents, SDR, applets …), has obvious associated security issues.
- Main issues are:
  - Integrity of code itself;
  - Authentication of origin of code;
  - Authorisation of code.
- Other issues:
  - Code confidentiality;
  - Intellectual property issues.

3. Use of trusted computing

- We conclude this talk by mentioning a few areas where trusted computing technology might help to address some of the identified mobile security and privacy issues.

Trusted computing I

- Trusted computing is hardware-based functionality in a computing platform that:
- (a) provides a hardware (cryptographic) basis of trust for system boot, integrity verification of applications, etc.
- (b) provides a means for an external third party to verify the current state of a platform.

Trusted computing II

- Considerable resources are currently being devoted to both hardware and software to support trusted computing (tc) technology.
- A number of vendors have produced TPMs (Trusted Platform Modules) conformant to the TCG (Trusted Computing Group) specifications.
- PCs incorporating TPMs are now widely available.
- Meanwhile operating system vendors and providers, including Microsoft and the open source community are working on operating systems exploiting this functionality.

Applications of tc

- There seem almost endless possible ways in which tc technology could be used to improve application security and privacy.
- The mobile world seems a natural environment in which some of the tc features can be exploited.

TC based stable identities

- A major problem in scenarios lacking security infrastructure is Sybil problem (entity claims multiple addresses) – e.g. in p2p and ad hoc settings.
- Trusted computing may be able to help by using the DAA protocol in a way which enables all actions of a particular platform to be linked, while not revealing true identity of that platform.

TC based trusted download I

- Problem arises in the context of broadcast to a mobile device.
- The established standard security techniques, e.g. involving use of broadcaster-specific smart cards in set-top boxes, is not really appropriate for a mobile model.

TC based trusted download II

- Use of a mobile platform for receiving broadcast content seems to mandate a software solution.
- However, the conditional access application needs to be protected.
- Conventional operating systems cannot provide the needed protection without a hardware token.
- Trusted computing can provide all the necessary guarantees.

TC based identity management I

- A range of different single sign-on (SSO) technologies exist.
- In a *true SSO* system, a user authenticates once to an Identity Service Provider (ISP), and this ISP then vouches for the identity of the user to multiple Service Providers (SPs).

TC based identity management II

- Clearly, the SP must trust the ISP to tell the truth about who has been authenticated and how.
- Typically this means that the ISP must be a networked entity remote to the user.
- The use of tc technology enables the ISP to be implemented on the user platform, in such a way that the SP can verify its trustworthiness.

TC based PI management I

- A growing number of possibilities now exist for Internet SPs to offer services tailored to end users.
- However, this possibility also represents a privacy threat, since the SP will typically need to know potentially privacy-breaching information about an end user in order to provide the tailored service.
- One key example (relevant in a mobile context) is the use of location information.

TC based PI management II

- How does owner of PI prevent it being disseminated and used in unauthorised ways?
- It is possible that tc can help.
- The holder of PI, and associated policy information (e.g. defining user preferences), can use tc functionality to check out the platform requesting PI, before sending it.
- This check could involve verifying the type of recipient platform and the identity of the receiving application.

TC-based co-operation enforcement

- The support of MANETs typically requires co-operation by the nodes, e.g. to support routing.
- Malicious users may replace their network software with a ‘selfish’ version, e.g. to save battery power.
- TC could help guarantee that a network element is running the ‘correct’ software, and hence will not behave selfishly.
- (Of course, this requires the communications hardware to be part of the TC subsystem.)

TC-based type approval

- The spread of computers everywhere (cars, fridges, toasters, …) gives rise to major problems regarding safety.
- For example, a car owner could replace the engine management software to radically increase engine power. See also SDR.
- Not only will this potentially wreck the engine, it may also be a major safety problem, since the brakes/suspension won't match performance.
- Traditional solution is a closed environment which will only run authorised software – however the trend is to open platforms everywhere, and TC may help give back control.