Security and trust for ubiquitous communication

S. Creese
Systems Assurance Group,
QinetiQ, Malvern Technology Centre, UK

G. M. Reed
International Institute for Software Technology
United Nations University
Macau

A. W. Roscoe and J. W. Sanders
Programming Research Group
University of Oxford
Oxford, UK

Prepared for the ITU WSIS Thematic Meeting on Cybersecurity
28 June – 1 July, 2005, Geneva Switzerland

June 15, 2005

Abstract

In this paper we report on an ethical approach to security and trust in contemporary computing and communication systems and we discuss some of its specific detailed consequences. The approach is that of *the principle of distribution*, in which control resides as much as possible with the individual rather than in centralised agents. The consequences on which we elaborate here include:

the importance of establishing entirely distributed protocols, in particular for the secure and trusted dynamic networking of mobile ICT devices;

that in view of their pervasive popularity, cellphones be subject to the same stringent criteria of security and trust as networked computers;

the promotion of open-source software to bridge the digital divide and empower developing nations in configuring accepted software in their own languages and to their own needs.

We report on the United Kingdom's Department of Trade and Industry's FORWARD program concerning security and trust in human-centric systems and on the United Nations University's International Institute for Software Technology's recent *Open Computing Initiative*, and end by raising some pertinent questions.

Introduction

1. We live in an age in which information and communications technologies span the globe, providing users with mobile and real-time access to information, services and each other. Increasingly the services offered are becoming not mere luxury but an established part of our everyday lives; a typical example is provided by the growing importance of e-services like e-government.[1] The resulting structure goes under a multitude of names;[2] here we refer to *ubiquitous communication* in the *comsphere*. By using 'ubiquitous communication' we mean to emphasise the importance of both synchronous and asynchronous communications and the growing mobility of the devices; and by using 'comsphere' (rather than the more accepted 'cyberspace') we mean to emphasise the difference that ubiquitous communication brings to the internet (e.g. the effect that mobility has on the resulting communications protocols). For this paper concerns precisely that distinction.

2. The twin features of globality and mobility provide distinct opportunities, but also reveal distinct difficulties. The former enables a global distribution of resources, but regardless of boundaries and perhaps therefore of propriety; the latter empowers users in remote or transient locations, but with an increased risk of insecurity. Means are needed to increase globality by increasing the penetration of ubiquitous communication amongst developing nations and to make the comsphere more secure. Those are the points addressed in this paper.

3. We begin by proposing a general but novel ethical principle, *the principle of distribution*, to facilitate the high-level discussion that ought to precede the more low-level concerns of technology, protocols and standardisation. In the present paper we concentrate on its application to the two points mentioned in the previous paragraph: security in the comsphere and making software more openly available, in spite of their apparent incompatibility.

4. At present the multifarious applications of ubiquitous communication remain largely untapped due partly to the increased opportunity that ubiquity in general, and mobility in particular, offer for malevolence. It appears vital that mobile users be able to generate spontaneously a secure network. That we consider in the middle section of the paper.

5. Finally we address the issue of making open source software available particularly to developing nations. The productivity and management processes appropriate to such novel modes of production yield unusual consequences for the assurance that open source software meets its requirements: it appears to be very difficult to certify such software. We are thus left with a divide between freely-available, reconfigurable (open source) software that is potentially of huge benefit in developing countries but for which authentication is difficult, and verified authenticated software that is necessary in a growing number of secure applications. Evidently both types of software are vital.

The principle of distribution

6. In this section we summarise the principle of distribution.[3] Its proposers have put the view that the standard normative ethical principles, which are the only apparatus provided by Ethics for use in the various fields of Applied Ethics (including Computer Ethics), are incomplete, having been proposed and developed to enlighten the individual in making moral decisions. In situations involving a single user of Information and Communication Technology they remain vital, of course. But in situations in which that technology manifests itself in a distributed system, with no central agency for coordination, they have been found to be inadequate.[4] That is scarcely surprising since distributed systems are comprehensively more complex than centralised systems, in exactly the same way that societies are comprehensively more complex than individuals.

7. The main concern with centralised control is that it is *fragile* in the sense that if the central agent becomes corrupted or fails then recovery of the entire system is extremely difficult. There is also a concern of *inefficiency*: if each individual in the system has to coordinate its activities with the central agent then many communications may be required and bottlenecks may cripple the system. An advantage, however, is that it is often conceptually simpler to design centralised systems. Recent examples in which distributed control has played an essential rôle are (a) the use of cellphones in responding to the Tsunami disaster and in organising demonstrations (in the face of centralised resistance),[5] and (b) in Pentagon defense.[6]

8. Distributed systems, and the notions on which they are based, have been studied in considerable depth in Computer Science.[7] For present purposes the idea is conveyed by comparing the games of soccer and baseball. The former might be said to be more distributed because no central agent is responsible for a team's play from moment to moment: the ball is passed between the distributed players following no centralised 'algorithm' but according to decisions made 'locally' by individual players, in spite of the 'global' aim of scoring goals. Indeed therein lies much of the interest of the game: how can such local decisions reach a globally desirable event? (Observe that the use of plays-in-a-down in American football imposes partial centralisation on such distribution.) By comparison in baseball there is far less scope for distributed decision-making: the game evolves on the basis of centralised decisions (except for double plays, and the routine decision by a fielder where to return the ball and a runner on a base whether or not to run for the next base, and how to do so).[8]

9. According to *the principle of distribution*, in a distributed system control should arise from the individual users (rather than being imposed centrally). Of course like any normative principle, the principle of distribution does not apply unequivocally and when it does apply it seldom provides the whole answer. It is a guiding principle to be used in conjunction with others in resolving complex issues, the most difficult of which pertain to control. It has implications for the design of protocols, the management of software development, education and policy. Let us consider a typical but simple example.

10. Many families find themselves confronted with the problem of what access to allow their children to the web.[9] A centralised or 'top down' solution would involve control

of undesirable sites. But one difficulty with that is: who has the right to make that choice for all, particularly in the context of the web? By comparison the principle of distribution leads us to consider 'bottom up' solutions, empowering individual homes: each household could filter access to the web using software chosen and configured by the guardians of the household.[10] (Ideally, free open-source software would be available online.) A similar solution applies to the screening of spam: each user can decide what he or she regards as spam and configure a filter program accordingly.

11. The dynamic bottom-up imposition of control takes time and is not appropriate in every situation. There are some situations in which a distributed solution does not exist.[11] It is, for example, unclear what is the right level at which to control malicious minority groups. The most surprisingly successful distributed protocols from Computer Science are perhaps those that use coin tossing to break symmetry. A primary example is Rabin's distributed algorithm for coordinating choice between two alternatives,[12] which can be understood like this.

12. A bus-load of tourists arrives in a new city and is to decide, by the end of the day, at which of two places to meet: inside a certain church or inside a certain hotel. There is no central agent (like a tour guide) and they may not communicate as a group: the tourists function as members of a distributed system. Rabin's algorithm shows that merely with a noticeboard at each location and a coin (to toss) for each tourist, by alternatively visiting each location and following a certain rule the tourists will, in a small number (with high probability) of visits, all end up by choosing the same location. A centralised goal has been reached on the basis of distributed decisions. Thus the principle of distribution is important in the design of realistic, efficient protocols in systems comprising many components.

13. Returning to cyberspace, the principle of distribution suggests:

    where possible, protocols for coordination of distributed users should be based on distributed control (to which we turn in the next section);

    individual users should be empowered by having access to the code on their devices. In particular open source software should be available online for routine (i.e. non-secure) applications, to help bridge the digital divide and to enable users in a variety of contexts to have access to accepted software and configure it for their own linguistic and social needs (we turn to this in the last section of the paper);

    where possible each device should be equipped with encryption to ensure secure communication and data storage;

    that e-services (e-business, e-government and so on) be freely available on as great a variety of devices as feasible.

The centralised alternatives allow: governmental monitoring of internet servers and text messaging; governmental intrusion; laws prohibiting encryption; and a software monopoly without open source. Naturally in practice some hybrid is to be expected. The most difficult issues involve: control during times of national disaster or invasion; undesirable information being available; anti-social behaviour; and concerning

economic responsibility. In the remainder of this paper we consider the first two of those proposals in greater detail.

Human-centric computing and FORWARD

14. Turning now to security in the comsphere, we adapt Kizza's definition[13] in the light of Schneier[14] and interpret *security* to consist of the process of maintaining:

> *confidentiality*: information is available only to those authorised to have it;
>
> *integrity*: data may be manipulated only by those authorised to do so;
>
> *availability*: information systems are accessible by all those authorised to access them.

Observe that the notion of *authorised access* underpins each requirement: access is permitted only if authority has been validated. Thus we concentrate on authorised access.

15. Centralised implementations ensuring authorised access (and hence also the requirements for security) are straightforward and rely on maintaining a central trusted list which is consulted to validate authentication. But in line with the principle of distribution it is preferable to use instead distributed authorisation (if possible).

16. Imagine that a group of you, not necessarily previously known to each other, meet (perhaps it is parents' night at the local school) and wish to form—spontaneously and in real time—a network with your wireless PDAs and cellphones. You cannot assume that your devices have unique identifiers or that any such identifiers are known in advance; and of course you wish to ensure that the network is established in a distributed manner, contains only those devices you want it to contain and that messages sent between you are secure to the network. You must assume, naturally, that none of you is malevolent. It is perhaps not obvious that those requirements can be met; but Creese *et al.* have provided and verified a protocol[15,16] which meets them. Its verification is achieved by strengthening the accepted model of security to take account of the flexibility evident in the comsphere; the formalism used is that of automated Communicating Sequential Processes.[17]

17. Traditionally it has been *identities* (of either users or devices) that are authenticated. But in the context of the comsphere it has been argued by Creese *et al.*[18,19] that it is *attributes*, and not identities, that must be authorised. Attributes include a device's location, name, manufacturer, internal state, service history and so on. Attributes appropriate to a given situation must be authenticated, and must be chosen to provide assurance not only about which devices are interacting but also about what they can do. This is an area in which more work is required.

18. The work reported in this section forms part of the FORWARD programme,[20] begun in January 2003 under the United Kingdom's Department of Trade and Industry's initiative into *Next Wave Technologies*. Part of the thrust of that programme has been the use of ubiquitous communication and computation to support human-centric

goals, like providing information in a form and at a time that is appropriate to the human user, and exploiting the human user's senses to complement digital bandwidth.

Open source

19. In this section we view software in the light of the principle of distribution. Commercial 'shrink wrapped' software may be seen as the result of a centralised process: the producer retains all rights and whilst allowing the user to use the code, does not provide direct access to it. The user is thus completely at a loss to modify the code in any way. By comparision, open source software may be seen as the result of a distributed process: it is typically available freely over the web and the user may take a copy to which he or she than has complete access. The differences between the two processes—the cathedral versus the bazaar—have been graphically documented.[21] The resulting difference is important, because having access to the source enables software to be adapted to its context, for example so that an interface appears with locally-appropriate features (at the very least linguistic). It also promotes local software productivity and so, eventually, promotes commerce. Perhaps it will one day produce a third-world Bill Gates!

20. But of interest to us here is the process underlying open source. In the standard model of software production, software is produced with some (varying, depending on use and style of software) degree of assurance that it meets its requirements. The extreme case is formally specified and verified code (like the protocols reported in the previous section). But of open source, what guarantees are there that a module downloaded from the web meets its requirements; and what protection is there against malevolent contributors to an open source project?

21. The response is to appreciate that a different model is involved. The production of open source, typical of an example of distributed control, is managed dynamically by feedback with some degree of conformance but also with attrition. Important, kernel, code is checked before release by one of a small number of agreed individuals. For less critical software, poor code suffers an 'evolutionary disadvantage' and is gradually superseded. This may seem strange from the traditional viewpoint based on the concern that even a single bug may lead to disaster. The conclusion, however, is simple. Open source and fully authenticated code lie at opposite ends of a spectrum, the whole range of which has a place in the comsphere. Fly-by-wire software, for example, would traditionally be produced by a more centralised process; uncritical applications software could be open source and so produced by a more distributed process. There remains the difficult issue of how much trust to place in any copy of a piece of software, whether downloaded or on disk, regardless of the claims that are made of it; but that is a topic of current research. We highlight the case of open source as being particularly important.

22. At the United Nations University's International Institute for Software Technology (UNU-IIST) in Macau, an *Open Computing Initiative* has recently been launched. The idea is to train representatives from third-world nations in the development of open source, thereby at once expanding the applications available in open source and

empowering third world programmers. Together with the fact that Negroponte's $100 laptop[22] will contain only open source software, we can expect a swing in the accepted style of software, from almost entirely centralised, commercial software to a more balanced hybrid of the two styles.

Conclusion

23. One of the specific WSIS plans of action is to:

> d) Invite relevant stakeholders, especially the academia, to continue research on ethical dimensions of ICTs.[23]

The principle of distributed ethics proposed recently and used here provides a response. It provides terminology with which to discuss principles as well as technical designs and is therefore appropriate for use in the range of endeavours from policy to standards. For more details on its application and its analysis as an ethical principle, we refer to the paper mentioned in Note 3.

24. We have argued that it is essential for the beneficial aspects of ubiquitous communication to function in the comsphere unchecked by malevolence and in a human-centric way. Accordingly we have stressed the importance of provably secure protocols for establishing spontaneous secure—possibly mobile—networks with the minimum of assumptions on the constituent devices. In this way we have addressed one of the WSIS principles:

> Within this global culture of cyber-security, it is important to enhance security and ensure the protection of data and privacy, while enhancing access and trade.[24]

In summarising the UK's FORWARD project we hope to have exemplified the WSIS principle:

> Many of the building blocks of the Information Society are the result of scientific and technical advances made possible by the sharing of research results.[25]

and to support the plan to:

> g) Share good practices in the field of information security and network security and encourage their use by all parties concerned.[26]

25. Open source software seems destined to play a vital part in empowering developing nations with ICT. We have strongly supported it, but raised some questions concerning its authentication and security and summarised the current model in which they might be appreciated if not resolved. Further work is needed here in order to conform to the WSIS plans of action:

> o) ... promote technologies and R&D programmes in ... a variety of software models, including proprietary, open source software and free software ...[27]

i) to encourage the development of content and to put in place technical conditions in order to facilitate the presence and use of all world languages on the Internet;[28]

26. Further consequences of the principle of distribution that we have noted in passing are:

in addition to its centralised, legally-enforced control (as discussed by others elsewhere), spam might be managed by use of filters (ideally open source[29]) residing on individual machines and configured by the individual to take into account his or her preferences;

Negroponte's $100 laptop might contain facilities for encryption and decryption of all communications, to avoid misuse by an oppressive regime.

The penalty paid for the former suggestion is some effort by the individual; for the latter the penalty is the far more serious problem of misuse by malicious minority groups.

27. We are left with many questions. Are there further distributed algorithms to secure the comsphere? How can control be achieved in a distributed and dynamic manner, perhaps in times of crisis? How can software (in particular open source) be assured to meet the claims made of its behaviour (like functionality or security)? How can distributed systems be developed and managed from partially centralised components? How can open source production be managed in a partially centralised project? It is hoped, in conclusion, that the principle of distribution will be useful in discussing policy as well as more specific matters of design, and that its consequences highlighted in this paper will add to the discussion at the Thematic Meeting on Cybersecurity, even if they do so by raising more questions than they answer.

Notes

1. See for example the international survey *Global Survey of e-Government*, by A. Ojo, T. Janowski and E. Estevez. e-Macau task 2 report, March 2005. To quote a specific instance, the aim of the recent u-Japan project is to make 80% of citizens feel comfortable with ICT, and to appreciate its rôle in resolving issues, by the year 2010. See `http://www.nri.co.jp/english/opinion/papers/2003/np200366.html`.
2. By comparison with comsphere, *cyberspace* is usually interpreted as comprising networked, static, users. It is vital for progress that mobility be acknowledged and catered for, particularly in the context of security, as we shall see in this paper. Just a few alternative terms for the activity of computing on the comsphere are:

   MOBILE COMPUTING: *IEEE Transactions on Mobile Computing*, founded 2002;

   PERVASIVE COMPUTING: *IEEE Pervasive Computing: Mobile and Ubiquitous Systems*, founded 2002;

   UBIQUITOUS COMPUTING: M. Weiser, "Hot Topics: Ubiquitous Computing", *IEEE Computer*, October 1993;

   PERSONAL COMPUTING: a term apparently coined by IBM and now interpreted more generally to mean individual local access to information facilities;

UBIQUITOUS NETWORK SOCIETIES: for the International Telecommunication Union's Workshop on Ubiquitous Network Societies, for example, see `http://www.itu.int/ubiquitous`;

UBIQUITOUS COMMUNICATION itself, interpreted to describe wearable systems: The UbiCom project, the Faculty of Information Technology and Systems at the University of Delft, led by R. L. Lagendijk.

3. The the principle of distribution was introduced as 'the principle of distributed ethics' in 'Ethical principles for secure ubiquitous communication', G. M. Reed and J. W. Sanders, draft May 25, 2005, written in response to the proposal 'Ethical strategies for human security', by Elisabeth Porter (research director for INCORE, the centre for International Conflict Resolution, a joint initiative between the United Nations University and the University of Ulster), which was circulated following CONDIR 29, 4–5 April, 2005, Bonn.

4. See the motivating paper 'What is computer ethics?', J. Moor. In T. W. Bynum (Ed.), *Computers & Ethics*, Blackwell, pp. 266–275, 1985. For treatments at the undergraduate level, see for example *A Gift of Fire: social, legal and ethical issues in computing*, S. Baase. Prentic-Hall International, 1997, and *Ethics and Technology: Ethical Issues in an Age of Information and Communication Technology*, H. T. Tavani. John Wiley and Sons, 2004.

5. See the articles by J. Yardley, April 25, 2005. Available at: `http://www.nytimes.com/2005/04/25/international/asia/25china.html?pagewanted=1` and Choe Sang-Hun, International Herald Tribune, 9 May, 2005.

6. For an interesting account of how a genuinely distributed system (investing as much control as possible in its disributed components) averted disaster when American Airlines flight 77 hit the Pentagon on September 11, 2001, see the article by R. Needleman at `http://www.microsoft.com/business/executivecircle/content/page.aspx?cID=979&subcatID=1`

> ... the system also remained functional even though a large part of it had been destroyed. ... In addition to playing well in large complex systems, they are able to autonomously perform actions that previously required a connection to a central control system.

7. The field is so young that many of the important textbooks study the topic using their own notation, which unfortunately makes them relatively inaccessible. A quite general textbook is *Distributed Systems: Concepts and Design*, G. Coulouris, J. Dollimore and T. Kindberg. Third edition, Addison-Wesley, 2001, whilst a slightly more representative text is *Distributed Computing: Fundamentals, simulations and advanced topics*, H. Attiya and J. Welch. McGraw Hill, 1998.

8. It is interesting to note the effect a malicious team member would have in each style of game. In the centralised game of baseball, were the pitcher or catcher in collusion with the opposition the result would be disasterous. However in the distributed game of soccer, a malevolent team member would be gradually marginalised (the most difficult case being the goalie, although defenders can to some extent compensate). This demonstrated the fragility of centralised systems mentioned in the previous paragraph.

9. "Two-point-five million use [America Online]. That's like a city. Parents wouldn't let their kids go wandering in a city of 2.5 million people without them, or without knowing what they're going to be doing." Pam McGraw, 1995; see `http://www.cybertoday.com/v1n4/runaway.html`.

10. See `http://safety.ngfl.gov.uk/?sec=9&cat=99&clear=y`. Relevant commercial programs include CyberSitter, SurfWatch and NetNanny.

11. For example if each user is to behave deterministically and identically then if they all start in the same state, no matter what communications they exchange and what internal decisions they reach, because they behave identically their subsequent states will remain identical. So they will be unable to reach a state in which one of them differs from the others.

12. 'The choice-coordination problem', M. O. Rabin. *Acta Informatica*, **17**(2):121–134, 1982.

13. *Ethical and Social Issues in the Information Age*, J. M. Kizza. Springer Verlag, 1998.

14. *Secrets and Lies: Digital Security in a Networked World*, B. Schneier. John Wiley and Sons, 2000. Schneier's point is that security is a dynamic process rather than a static product.

15. 'The attacker in ubiquitous computing environments: Formalising the threat model', S. J. Creese, M. H. Goldsmith, A. W. Roscoe and I. Zakiuddin. In *Formal Aspects of Security and Trust*, Pisa, Italy, September 2003. IIT-CNR Technical Report, edited by T. Dimitrakos and F. Martinelli, 2003.

16. 'Exploiting empirical engagement in authentication protocol design', Sadie Creese, Michael Goldsmith, Richard Harrison, Bill Roscoe, Paul Whittaker and Irfan Zakiuddin. In D. Hutter and M. Ullmann (editors), *SPC 2005*, Springer LNCS **3450**, pp. 119–133, 2005. Since that paper has appeared the authors have extended the same technique to networks including mobile phones.

17. *The Modelling and Analysis of Security Protocols: the CSP Approach*, P. Y. A. Ryan, S. A. Schneider, M. H. Goldsmith, G. Lowe and A. W. Roscoe. Addison-Wesley, 2001.

18. 'Authentication in pervasive computing', S. Creese, M. H. Goldsmith, Bill Roscoe and Irfan Zakiuddin. In D. Hutter and M. Ullmann (editors), *First International Conference on Security in Pervasive Computing*, Boppard. Springer LNCS, 2003.

19. 'Research directions for trust and security in human-centric computing', Sadie Creese, Michael Goldsmith, Bill Roscoe and Irfan Zakiuddin, 2005. At:

20. See `www.forward-project.org.uk`.

21. For a graphic exposition of the different business models appropriate to commercial software and open source, see *The Cathedral and the Bazzar*, E. S. Raymond. O'Reilly, 2001; in particular the article after which the book is titled, pp. 19–63, and 'The magic cauldron', pp 113–166.

22. See `http://laptop.media.mit.edu`. A similar, but established and successful project, is the Jhai Foundation's PC used in particular to provide internet access

to villages in Laos without electricity; see
http://www.jhai.org/jhai_remoteIT.htm.

23. 'WSIS Plan of Action', document WSIS-03/GENEVA/DOC/5-E, 12 December 2003. Paragraph 25(d).
24. 'WSIS Declaration of Principles', document WSIS-03/GENEVA/DOC/4-E, 12 December 2003. Paragraph 35.
25. *loc. cit.* paragraph 7.
26. 'WSIS Plan of Action', document WSIS-03/GENEVA/DOC/5-E, 12 December 2003. Paragraph 12(g).
27. *loc. cit.* paragraph 23(o).
28. *loc. cit.* paragraph 6(i).
29. See PopFile, for example, at http://getpopfile.org.