The Butt of the Iceberg: Hidden Security Problems of Ubiquitous Systems

Frank Stajano
University of Cambridge, Laboratory for Communication Engineering
http://www-lce.eng.cam.ac.uk/~fms27/

Jon Crowcroft
University of Cambridge, Computer Laboratory
http://www.cl.cam.ac.uk/~jac22/

Abstract

In Kurt Vonnegut's eponymous novel, there is a moment when an adult demonstrates Cat's Cradle, the art of making patterns with a loop of string held between two hands, to a small child. On dropping the string, the adult says "Look: no cat, no cradle". The child is naturally dismayed.

In a sense, ubiquitous systems are as ephemeral as the Cat's Cradle. The infrastructure to support them is as subtle, and complex, and the trust we have in them as fragile. This paper reveals the hidden problems in store for security in ubiquitous systems. A crucial aspect we wish to stress is that there is a phase change as one moves from classical distributed computing into this new environment: the mere shift in quantity of components is enough to lead to qualitative changes in aspects of security. Moreover, there are inherent qualitative shifts.

Keywords: ubiquitous computing, security, privacy

1. The Tip of the Iceberg

It is inevitable that the physical manifestations of personal computers will vanish into the surrounding infrastructure and into the very fabric of buildings, clothes, vehicles and so forth. As this happens, computer scientists will need to come to terms with systems that interact in an ad hoc manner with their users, as well as with each other. Concepts of input, output, persistence of information and computation will change. The notion of ownership of information and computation will drastically evolve, along a trajectory that is only partially visible today.

The expression "ubiquitous computing", now often abbreviated to "ubicomp", was made popular by the late Mark Weiser in his well-known 1991 *Scientific American* article [12]. In that early work he candidly admitted that systems implementing his vision would evolve in ways he could not possibly foresee:

> Neither an explication of the principles of ubiquitous computing nor a list of the technologies involved really gives a sense of what it would be like to live in a world full of invisible widgets. To extrapolate from today's rudimentary fragments of embodied virtuality resembles an attempt to predict the publication of *Finnegan's Wake* after just having invented writing on clay tablets. Nevertheless the effort is probably worthwhile.

In 2003, over a decade later, ubicomp is irrevocably happening. On the computing side, even non-technical people unwittingly interact on a regular basis with invisible microprocessors embedded in vehicles, buildings and entertainment systems. On the communication side, cellular telephony has been growing steadily at an even faster rate than computing; and wireless networking for mobile computing is becoming a commodity item: the availability of 802.11 connectivity at conferences, as happened for VGA projectors a few years back, is soon graduating from a novelty feature to something that attendees naturally expect. Many consequences of ubicomp that were not believable or not obvious back then are now in plain view. But this is still only the tip of the iceberg—more developments will come along and surprise us.

In this paper we explore some of the outstanding long-term challenges for ubicomp, with special attention to the systemic aspects. Embedded computing and wireless networking are no longer in their infancy, but what is still missing is a synergy of the parts: bringing together all these hidden processing nodes into a coherent entity worthy of being called a *system*. Once this is achieved, the complex system will exhibit interesting properties, of which *security* will have special importance. Security is a paradigmatic example of a system property—one that can only be achieved holistically. In these pages we aim to understand the challenges ahead by imagining how ubiquitous systems may develop.

We have been writing on clay tablets for a little while now; perhaps we are already scribbling on papyrus. While it may still be early to envisage *Finnegan's Wake*, for a field that evolves at such speed there is merit in cranking up the imagination dial rather than considering only small incremental developments, even if the result borders on the outrageous. In the next section we are going to do just that.

2. Application Scenarios for Ubiquitous Systems

Where could ubiquitous systems take us, in the course of the next decade, from the viewpoint of technological feasibility? Let's have a little brainstorming about what might happen, without worrying too much for the moment about whether it actually will.

2.1 Always-on Health-Care

Imagine the entire population of your country being permanently wired for systematic monitoring of various metrics such as heartbeat, skin conductivity, blood sugar, etc. Imagine this data being locally and remotely logged in a secure way via wireless communication. In the short term this is done mainly for people at risk and the data is "phoned in" periodically. In the longer term, monitoring covers everyone, on a 24×7 schedule. One possible justification for this development is long term medical research.

The next stage is to introduce intervention. Firstly, via remote diagnostics and human advice; later, via autonomic responses (e.g. triggering defibrillators for cardiac patients undergoing unattended attack; appropriate other responses for epileptics, diabetics and so on).

The third stage is to synchronize information (telemetry, long term data and local data) between patient, hospital and the paramedics of an arriving emergency service. It may sound optimistic to imagine that the synchronization and intervention will be achieved without problems. However, in any case, the risk levels could be provable properties of the system: an automatically derived explanation would be provided in case of failure, such as "it ran the defibrillation, despite there being a 12% risk that this was not heart failure, because there was a 0.003% risk of this doing harm and a 48% chance of it being beneficial".

It has been observed that in fact deriving this type of explanation is a very hard problem, and that it is all too easy to fool the user into thinking that they have a deeper understanding of a complex system (e.g. the heart) than really possible.

2.2 Surveillance and Evidence-Gathering

Nowadays, in our society, a large proportion of people have mobile phones, enjoy Internet access and exchange electronic messages with greater frequency

than they write snail mail. Many commercial transactions, and certainly most of the higher-value ones, are paid for with plastic rather than cash. Furthermore, especially in the UK, most public and many private spaces are under video surveillance. As a consequence, each one of us leaves behind a conspicuous digital trail [5] through which our past movements and actions can be reconstructed by third parties who were not involved in them.

Various agencies will be keen to use more and more of this data in ways that enhance public safety and offer better (in the sense of accurate and provable) evidence for criminal charges. Ideally, this would create disincentives for criminal behaviour without intruding on the privacy and other human rights of the general population.

Two real world examples suggest directions for this.

Locating Mobile Phones. In recent test cases [1, 2], mobile phone call records (which can establish the phone's location quite accurately, to about 100 metres) have been used in defence as "evidence" of a person's absence from a scene.

Strictly speaking, all that such records can prove is the location of the *phone*, not of its owner (short of recording the call and recognizing the voices of the correspondents; but this would require prior warrant to tap the call). For this reason, they cannot be used by the prosecution as evidence of the presence of the accused party at the scene; whereas, in the opposite sense, a defendant is automatically granted the benefit of the doubt, which the hint offered by the location of the phone can corroborate.

The increasing pervasiveness of video surveillance will certainly offer further opportunities to build up "evidence" by combining different kinds of such hints.

Cameras Reading Car Number Plates. In February 2003, the city of London launched a scheme to charge motorists in order to reduce congestion. Cameras along the boundary of the restricted zone check car registration plates automatically and issue fines to the drivers who have not paid the charge.

In the original design, cameras used to monitor vehicles would immediately discard the acquired combination of registration plate, location and time if the database listed the car as having already paid. Under such a policy, there would be no record of the movements of law-abiding people (or, more accurately, vehicles, since the system cannot tell who is driving).

Recently, though, the policy was changed: the police and other authorities will be granted routine access to the location data of law-abiding users, of course within the guarantees of the Data Protection Act.

This poses several threats. Given there are so many people with access to the police computers it is relatively likely that someone can be bribed or

blackmailed into giving over data. Drivers may then be embarrassed (e.g. when journalists report them "at their lovers' nest") or fired (when seen at the match instead of home sick) or killed (when terrorists learn the routes taken by their political targets to their offices and constituencies).

2.3 Zero Road Fatalities

More young people die on roads than from any disease. It is amazing that this is socially acceptable. It should be possible to build a system that reduces death for pedestrians, cyclists and drivers to zero. A low-tech solution is to put a 15 cm spike on the centre of the steering wheel—this provides an interesting incentive to drivers. Another is to require an attendant to walk in front of each car with a red flag.

Realistically, it is already possible to get cheaper insurance for sports cars if the system is fitted with a speed limiter. As other mandatory safety features, from seat belts to airbags, gain broad social acceptance, we may imagine the next step to be the automatic override of the vehicle's controls.

With sensors monitoring the presence of humans in the vicinity (on the sidewalk, or nearby and on a collision trajectory) and of obstacles in general, a control system could adjust the speed of the vehicle to a safe limit whenever someone is near. Incentives such as insurance discounts could be introduced to accelerate the deployment. And the system would not necessarily have to force ridiculously low speeds.

The interesting problems with such a system are clear when you consider what a driver would do—people would probably put their foot to the floor at all times. What then if the system fails?

Going further, it is not an enormous leap of the imagination from this to a car equipped with a navigation system, inter-vehicle communication, road traffic monitoring and fully automated driving. Could we then speak a destination, sit back and read the paper? At least in theory, with all cars smoothly cruising along at a constant safe speed without overtaking, this kind of system could not only improve safety but also offer the guaranteed arrival times of a conveyor belt, combining the reliability of a dedicated subway system with the comfort of a private vehicle.

3. The Shape of Ubicomp

Ubiquity of computing and communication poses major technical challenges across a broad range of computer science topics. We will require a coherent approach to networking, operating systems and to the various programming environments for applications, interfaces and user customization. A good Ubiquitous System must accommodate new structures such as combinations of unreliable components and yet appear more reliable and usable than today's PCs.

3.1 Quantitative Changes

The exponential growth predicted by Moore's Law has continued to hold for much longer than originally expected. A rough but essentially accurate summary of the performance evolution of a standard $1000 computer over the past two decades is that we went from "3 Ms" (MB, MHz, Mbps) to "3 Gs" (GB, GHz, Gbps). The trend continues towards the "Ts". This is now far more than enough for most sensible personal computing needs, especially given the ability to provide services within the network. (Except that we keep on inventing new ways to soak up memory, processor cycles and bandwidth—not to mention disk space.)

An apparent trend in computer science is that we will soon produce constant performance at a falling price on the same curve, leading to the facilities in a 2003 desktop being available for under $1 by 2013. There is evidence that this direction is already being followed to some extent, with PDAs, mobile cell phones and hand held gaming computers falling below the $100 mark. However a lot of this is achieved through a reduction in functionality.

Currently, smaller inexpensive computers have user interfaces of limited functionality (displays of lower resolution and smaller physical size, pens instead of keyboards etc.). By contrast, with 3D projection display, voice input and output, and other interaction modes such as gesture, one could consider a wholly different style of use. This would not be a computer for embedded systems work (as in today's phones or PDAs), or a pure tablet (a.k.a. "network computer"), but a fully featured device. The best description of the type of interface we envisage, for science fiction fans, is probably to be found in the 1953 Asimov novel *Second Foundation*, which introduces the Prime Radiant, with full fledged computing, communications, storage, etc. Such a computer may be worn, but it may well also be something that one places in large numbers around one's person and property, as well as in the broader, public environment.

This latter point is critical. When this approach is used, a merely quantitative change in the number of computing devices leads to a qualitative set of changes in the way one deals with information and computation. Key to this is the notion of ownership. The physical instantiation of computing is no longer there to allow one to have even the illusion of ownership and control.

At the same time, comparable computational power is becoming embedded in dedicated objects, as is already happening today, with the difference that ubiquitous networking would bring them together in a system as opposed to isolated parts. Let's briefly review the way this is happening.

3.2 Qualitative Changes

There are already many more computing devices embedded in the environment than associated with individuals in the form of PDAs or wearables. However, to date, these devices are largely independent (e.g. engine management systems), replacing the analog control systems of the past. There is great potential in connecting them into a pervasive information sensing and actuator system.

Already, connectivity such as Bluetooth allows the user to control several objects near their phone. Wireless LAN hotspots and GPRS allow the laptop or PDA to integrate into the world in many places, and to access remote devices. Low cost home area networks allow domestic appliances and sensors to be accessed and managed by a central computer or from outside the home.

Networking all of these components that were previously autonomous leads to a number of important consequences that we address next.

4. Challenges for Ubiquitous Systems

4.1 Control

Many synonyms have been coined to describe the vision of embedding computing and communication capabilities into everyday objects: we have so far adopted the Weiser locution of "ubiquitous computing" but the reader will have also heard these same ideas variously described as "proactive", "context-aware", "sentient", "calm", "ambient" and "pervasive" computing. This last term, "pervasive", while popular in some circles, is avoided by a number of researchers who prefer not to evoke any sinister overtones of wide-reaching domination of machines over mankind. We, instead, are keen to awaken these fears explicitly in order to help ensure that they won't have a chance to become reality.

At the technology level, the security issues posed by ubiquitous systems can be split into two main categories: the ones that can be solved with the traditional tools developed for distributed systems security, and the new ones that require original solutions. The latter group includes for example secure ad-hoc routing [6], stream authentication for sensor networks [7] and secure transient association [10].

At the highest level of abstraction, though, the application scenarios we presented in Section 2 highlight one major trade-off that ubicomp forces on us: the convenience of self-activating and proactive technology is inescapably intertwined with the risk of losing control. The most fervent proponents of the new way will be happy to delegate away to their car the low-level chore of driving, in the same way in which they gladly switched to automatic transmission

years ago; many others, though, are bound to remain sceptical. The trade-off is even more evident in the health-care example.

4.2 Ownership

The problem of control, or lack thereof, is compounded by that of ownership. In a world of systems as opposed to self-contained devices, the service we receive is the result of a synergy between many devices, not all of which will be owned by us.

In one of our ongoing experiments [11] we are building a "sentient car" that, among other features, is capable of seamless handover on the move between heterogeneous wireless networking technologies: it will switch from GSM to 802.11, without dropping connections, as it drives through a "Wi-Fi" hot spot. The connectivity provided by this hot spot, however, is made available by third parties under terms over which the car driver has no say (other than perhaps a boolean "accept or reject"); in this sense it is a rudimentary example of an external dependency. The ubicomp world will present us with many more such dependencies, so the concept of virtual ownership must be addressed.

The move from the mainframe to the PC was a move from central to personal ownership. The systems we envisage are a move towards fully cooperative virtual ownership. Notions of identity and provenance are critical. There are many challenging technical problems here, but also social value: disposable grids; affordable computing for the developing world; collaborative filtering; emergency service and disaster recovery support; location aware services and so on.

Some previous work [10] suggested a strategy through which devices could be securely "lent" to temporary owners. More work is needed, though, on higher level incentives, rewards and accounting.

Ubicomp extends even to "smart dust" or similar intelligent sensor net computing projects: here, a combination of systems co-operate to offer functions previously located in in one small device (personal communicator/digital assistant/games console/toaster). The dynamicity of these systems will be extreme. As users move amongst the computing milieu, they will autonomically create services as easily as one today borrows and uses physical tools such as knives and forks, whether at home, at a friend's or in a restaurant.

4.3 Privacy

The application scenarios of Section 2 also make it evident that ubicomp may become a serious threat to privacy.

At the technical level, some basic countermeasures may include anonymization and brokerage of queries through a layer of trusted middleware. For example, in the surveillance scenario, the middleware would disallow access to

the raw data and only permit queries according to a standard set of interfaces. These might enforce rate limitations (limiting the frequency with which a certain kind of question can be repeated) and semantic limitations (allowing for example "were you at location L at time T?" but not "where were you at time T?"). These safeguards would be provable properties of the middleware, which would be open to inspection and verification by the users or their appointed expert advisors. Preventing attackers from inferring higher level information from the allowed queries is, however, a hard theoretical problem [4]. On the practical side, instead, ensuring that the middleware cannot be bypassed to access the data directly is going to be equally arduous.

Going up one level, we ought to be cynically sceptical about the guarantees provided by policy assurances such as the ones that the above middleware layer would be charged with enforcing. The London congestion charging scheme mentioned above provides a neat example of a significant change in policy introduced after the deployment of the surveillance infrastructure.

As appropriately noted by Zimmermann [13] well before 9/11,

> while technology infrastructures tend to persist for generations, laws and policies can change overnight. Once a communications infrastructure optimized for surveillance becomes entrenched, a shift in political conditions may lead to abuse of this new-found power. Political conditions may shift with the election of a new government, or perhaps more abruptly from the bombing of a Federal building.

Responsible architects of ubiquitous systems will take this into account and examine in detail the ways in which their envisaged systems could be abused.

There is often a conflict of interests for researchers working on new and exciting technologies: being too lucid and honest about the dangers for society of one's new creation may hamper or block its development. It is therefore easy to self-justify the decision not to spend time and effort trying to limit the potentially dangerous capabilities of the invention by arguing that other less scrupulous competing researchers would otherwise be first to market, of course with the unrestricted version. It is also frequently argued that, except for extreme cases, it is not the technology itself that is inherently evil but only some of the uses to which it may be put. While these arguments have merit, it would be shortsighted to use them as excuses to avoid a critical assessment of the risks that the new technology might pose to society in case it were intentionally misused once deployed.

Let us revisit some of the problems associated with just one aspect of security, namely the normal non-technical notion of privacy, to see some of the effects of ubiquity. Spärck Jones [9] envisages the following properties of the technology that create novel pressures on the perception of privacy.

Permanence Information in a ubicomp world attains a persistence hitherto unforeseen.

Volume The range of types of data available in the infosphere is massively increased.

Neutrality The data may easily be divorced from associated meta-data (or real world context, such as location, ownership) and as such used in ways that are unexpected.

Accessibility The availability of data on a network is much wider.

Assembly Data may be combined and mined.

Remoteness Aliens may see data.

This adds up to another vivid representation of the fact that ubicomp is no longer just a quantitative change but a significant *qualitative* change, bringing about new issues, opportunities, problems and threats that simply didn't exist before.

5. Mistrust

As computers disappear into the environment they become less threatening to non-technical users but at the same time even more mystifying, especially when they don't work as expected. It is hard enough for a regular human being to understand why a regular computer misbehaves; it may be a lot harder when there appears to be no computer. Witness digital cameras (or cars!) that lock up and can't even be switched off with their "off" button—the only escape route being a hard reboot via removal of the battery pack.

Some current research [8] uses augmented reality to visualize the state of ubicomp devices that lack an obvious user interface. The effort needs to scale up to higher levels of abstractions too, in order to help non-technical users build correct and reliable mental models of the behaviour of the ubiquitous systems around them.

Earlier on we discussed a health-care system that could trace back and justify its choices—particularly the ones that later turn out to be inappropriate. More still needs to be done on this aspect, especially with regard to incorporating human intuitions about the expected behaviour. The more ubicomp systems behave counterintuitively and unpredictably, the less they will be trusted by their users.

6. Conclusions

We have examined some consequences for security for ubiquitous computing at the levels of systems and of users. The assertion is not that ubicomp is bad. Rather, we want to draw attention to the qualitative nature of the changes brought about by the quantity and type of data and processing available in such

an environment. We could picture a major crisis for the introduction of new services if awareness of this big picture is not taken on board. The types of reactions might fall into the following categories.

- Principle of unintended consequences
- Confounded expectations
- Surprise
- Frustration
- Rejection

These reactions might not be irrational.

7. Acknowledgements

Some of this material, including the three example scenarios, was originally developed by a brainstorming group sponsored by the BCS and UK CRC [3] in the context of the UK Grand Challenges in Computer Science initiative. We also acknowledge a useful discussion of novel problems in privacy with Karen Spärck Jones.

References

[1] BBC News. Evidence 'proves Hamiltons innocent', 2001. `http://news.bbc.co.uk/1/hi/uk/1488859.stm`.

[2] BBC News. At a glance: Damilola CPS report', 2002. `http://news.bbc.co.uk/1/hi/uk/2558777.stm`.

[3] Jon Crowcroft and George Coulouris. Scalable Ubiquitous Computing Systems or just Ubiquitous Systems. `http://umbriel.dcs.gla.ac.uk/NeSC/general/esi/events/Grand_Challenges/proposals/US.pdf`. (Offline, but cached in Google, as of 2003-06-17.).

[4] Dorothy Denning. *Cryptography and Data Security*. Addison-Wesley, 1982.

[5] Simson Garfinkel. *Database Nation*. O'Reilly, 2000.

[6] Yih-Chun Hu, Adrian Perrig, and David B. Johnson. Ariadne: a secure on-demand routing protocol for ad hoc networks. In *Proc. MOBICOM-02*, pages 12–23, September 2002.

[7] Adrian Perrig and J. D. Tygar. *Secure Broadcast Communication In Wired and Wireless Networks*. Kluwer, 2002.

[8] Kasim Rehman, Frank Stajano, and George Coulouris. Interfacing with the Invisible Computer. In *Proc. NordiCHI 2002*, October 2002.

[9] Karen Spärck Jones. Privacy: What's Different now?, October 2002. British Academy Conversazione, `http://www.cl.cam.ac.uk/users/ksj/privksjtext1.pdf`.

[10] Frank Stajano. *Security for Ubiquitous Computing*. John Wiley and Sons, February 2002.

[11] Pablo Vidales and Frank Stajano. The Sentient Car: Context-Aware Automotive Telematics. In *Proc. LBS-2002*, September 2002.

[12] Mark Weiser. The Computer for the Twenty-First Century. *Scientific American*, 265(3):94–104, September 1991.

[13] Philip R. Zimmermann. Testimony of Philip R. Zimmermann to the Subcommittee on Science, Technology, and Space of the US Senate Committee on Commerce, Science, and Transportation, 26 June 1996.