

Proximal Interactions: A Direct Manipulation Technique for Wireless Networking

Jun Rekimoto, Yuji Ayatsuka, Michimune Kohno, Hauro Oba

Interaction Laboratory

Sony Computer Science Laboratories, Inc.

3-14-13 Higashigotanda

Shinagawa-ku, Tokyo 141-0022, Japan

Phone: +81 3 5448 4380

Fax: +81 3 5448 4273

Mail: {rekimoto,aya,mkohno,oba}@csl.sony.co.jp

<http://www.csl.sony.co.jp/person/rekimoto.html>

Abstract As a number of networked digital devices are ubiquitously used, control of inter-device communications becomes a complicated task. People can no longer keep track of all the devices' addresses and name. Some devices (such as wireless headsets) have very limited information display capability, making it difficult to use normal GUI techniques (such as menu selection). This paper proposes a solution to these problems by introducing a nearfield communication channel based on radio-frequency identification (RFID) or infrared communication technologies. By using this nearfield communication channel in conjunction with a normal wireless network, people can establish wireless connections among nearby devices with intuitive actions such as directly pointing one device at another or putting two devices within close proximity of one another. Secure network communication is ensured by passing session key information over the nearfield channel. This paper presents the basic idea, network protocols, and several applications, including inter-device data transfer, universal remote commanders, and extensible mobile IP-phones.

Keywords: wireless networking, ubiquitous computing, mobile devices, service discovery

1 Introduction

As Mark Weiser suggests [14], our styles of using computers are radically shifting from PC-centric computing to "ubiquitous computing." In the ubiquitous computing environment, many different types of computers, or digital devices, enable communication via wired or wireless networks. The concept of "computer" is not limited to today's general purpose "personal computers," but also includes digital appliances such as TV sets, cellular phones, personal digital assistants, portable audio players, or even wristwatches. Users will perform tasks through the combination of several devices. They will purchase a digital movie using a cellular phone, and watch it on a TV at home. They will also take pictures with a digital camera, view them on a computer screen, and send them to other devices. In such environments, network connectivity (both wireless and wired) is of utmost importance. Network configurations are not stable, but are dynamically changing.

However, without an easy and intuitive User Interface, such environments could be very frustrating and

perhaps unsafe. A summary of the problems follows.

Identifying target devices:

Traditionally, networks were used for connecting remote devices. For this purpose, specifying a network address, such as the Internet address is inevitable. Recently, networks, especially wireless ones, have also been used for connecting nearby devices. These devices range from normal computers to various types of digital appliances, such as PDAs, cellular phones, or digital TVs. One of the expectations of future ubiquitous computing is that a user will always carry a mobile device and will use it for controlling nearby devices, and for transferring data from one device to another. In such environments, wireless network connections will be frequently established, and configurations will be dynamically changed, according to the user's location or intentions. Since most of the target devices are within the user's physical reach, requiring a user to enter a network address (such as an IP address) when connecting to these devices is not a good idea. It is difficult to assume that

users always remember all the IP addresses of all surrounding devices. Furthermore, many devices now rely on a dynamic IP address allocation mechanism, such as DHCP¹ or AutoIP [3], instead of static IP addresses. Then it becomes impossible to know beforehand the address corresponding to the specific device. From the user's point of view, there should be more direct and unmistakable ways to identify devices.

Several recent activities on network services try to provide a method for accessing network resources through more understandable names such as "Kate's PC" or a service name like "Printer in the copier room." However, maintaining these names and real addresses still requires considerable effort. Some digital devices, such as wireless headsets, do not provide a GUI interface for selecting target, thus a name-based selection is not always a good solution.

Security and authentication:

Another issue that becomes increasingly important is how to intuitively establish and control secure communication in a ubiquitous computing environment. Security policies based on the firewall are inadequate because it assumes static organization of computers (i.e., assumes computers to permanently connect to the fixed position). User authentication based on passwords is increasingly cumbersome because users may change the destination devices with a high frequency and entering a password every time becomes unrealistic.

These problems stem from the fact that traditional networks treat all the network communications the same. In actuality, though, accessibility is naturally separated by physical context. Personal computers located in personal offices are considered to be more secure than the ones in common terminal rooms. There are physical and social barriers that prohibit unknown outsiders from accessing personally owned computers. When leaving an office, we simply lock the office room instead of locking (logging out) the computer.

2 The Proximal Interaction Model

Our proposed user interface model addresses these problems by introducing another communication channel, called the *nearfield channel*, in addition to conventionally used normal wireless networks (here we refer to these as the *standard channel*). The nearfield channel is typically a range-limited wireless data transmission such as infrared beaming or radio-frequency identification (RFID) technologies (Figure 1). The nearfield

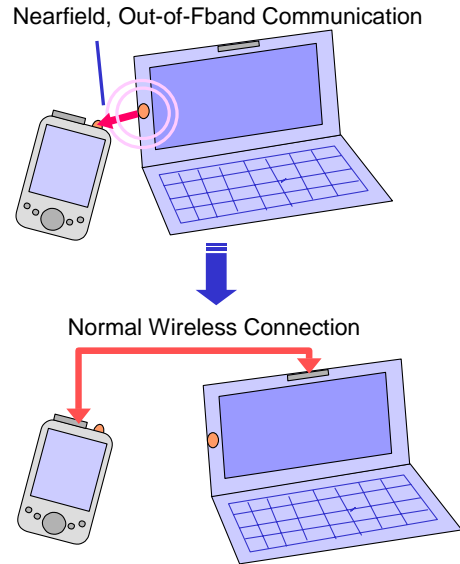


Figure 1: Proximal Interaction Model



Figure 2: Two typical operation styles of proximal interaction. (left: two devices within close proximity, right: beaming at target device)



Figure 3: An example of enhanced PDA. Two proximal interaction channels (RFID tag and infrared beaming) are available in combination with normal wireless LAN.

channel might be a complete bi-directional wireless channel, or a one-way data transmission (e.g., RFID tags). To set up a wireless communication between two devices over the standard channel, a user first puts two devices within proximity of each other, or pointing one device at the other (i.e., beaming). Then, the nearfield

¹ Dynamic Host Configuration Protocol

channel is used to transfer necessary information for setting up a connection, such as the device's address or a session key (Figure 2). Once the information is transmitted over the nearfield channel, devices establish a wireless connection over the standard channel. After that, these devices should have no restrictions within range of the nearfield channel.

By combining this nearfield channel and the standard wireless channel, such as IEEE 802.11a,b,g or Bluetooth, a wireless connection can be set up by performing a direct operation. During the operation, users are not required to enter any addresses or passwords. In this paper, we present several application systems based on this idea, as well as internal system architecture and network protocols.

3 Related Technologies

Our research described in this paper is an attempt to introduce the notion of "direct manipulation" [9] into the physical world. We are inspired by a series of previous research aiming at a similar goal. "Pick and Drop" [7] extends a popular "drag and drop" GUI technique and handles inter-computer data manipulations. MediaBlocks [11] uses a tagged physical block as a graspable data container. Both systems eliminate the necessity of identifying computers by addresses or names. Instead, more direct operations such as pointing can be used. These systems mainly concern one-shot data transmission, and continuous network connections are not explicitly supported. Our work also supports secure wireless connections, where these previous systems did not support security.

Our approach differs from short-range wireless technologies such as Bluetooth [2]. Bluetooth's communication area normally covers a ten-meter range and it is hard to distinguish a target device among other devices within this range. Our proposed method uses more user controllable methods such as RFID or infrared, and the Bluetooth connection could be established by the nearfield channel.

Although data transmission by infrared light has been used for a long time (such as IrDA – infrared data association [4]), because of several limitations (e.g., devices cannot move freely to keep a line-of-sight connection during data transmission), this technology devolves its position to wireless networks. Our proposed method regards infrared beaming as a trigger for a wireless connection, instead of using it as a primary data transfer channel.

Infrared beacons are widely used in museums for providing context information to mobile navigation de-

vices. The active badge system [12] deploys an infrared sensor in each office to detect people's location. The Cooltown project uses infrared transmitters that periodically emit location-related URLs[5]. There have also been a wide variety of systems that uses RFID tags to connect physical and virtual. For example, Want's system shows how everyday objects such as books, can be enhanced by attaching an RFID to them [13]. Other tagging technologies, such as those based on visual patterns, are also used to identify target devices [8]. While these systems use tags as simple identifiers, our system also combines communication with wireless networks.

Active research is being done on configuration-less networking where computers and digital devices can dynamically join the network. For example, IETF Zero-Conf [15] defines a set of internet protocols that support dynamic assignment of IP addresses without a central (i.e., DHCP) server, and supports discovery of resources by device name (e.g., "Printer in the copier room"). However, handling such nicknames for many devices requires human effort (i.e., who decides when the printer is moved to the meeting room?), and selecting a device name from the long menu list is also a cumbersome task, especially when the target device is in front of a user.

Ensuring secure wireless communication with intuitive user interfaces is also important. Despite a long history of secure network protocols, user interfaces for secure networks have not been well studied until recently. Stajano and Anderson introduced the "Resurrecting Duckling" security model [10], where devices establish a transient association by physical contact. The Bluetooth SIG also suggests physical connection as a method for defining association and passkey sharing between devices. Balfanz et al. propose the use of the out-of-band channel to exchange the session key information [1], which is similar to our (bi-directional nearfield channel) protocol. In this case, physical contact is not a prerequisite, but nearfield communication methods such as infrared beaming can be used for this purpose. We also designed a protocol to be used where only a unidirectional nearfield channel is available. We also developed various applications and user interface techniques which have not been well studied in past.

4 Applications

This section presents several examples for showing how the combination of the nearfield and the standard wireless channels are used in various situations.

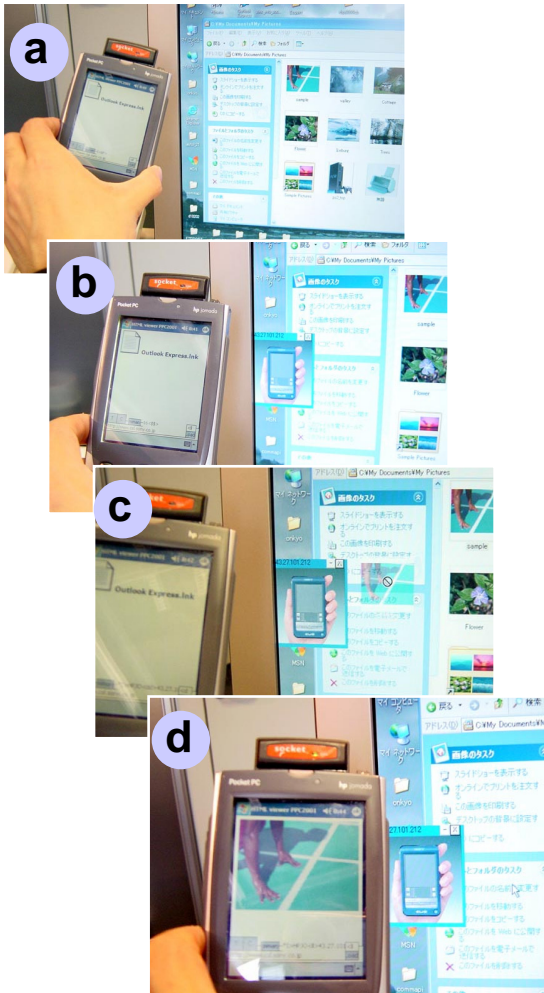


Figure 4: Ad hoc wireless connection data transmission using proximal interaction: (a) User attaches PDA to display side. (b) Wireless connection between PDA and computer starts, and a portal window appears on the display. (c) User drags an onscreen item (such as a file) to this window, (d) Data is transmitted through wireless network connection.

4.1 Setting up a wireless connection between devices

For configuring wireless peripheral devices, such as a wireless mouse or a wireless headset, a user is typically required to enter the device identification number before using it. Using the proximal interaction model, a user simply "attaches" a device to another device in order to make a wireless connection (Figures 4 and 5). When a PDA with an RFID tag (Figure 3) is attached to the other computer, the ID is recognized and a wireless connection for data transmission is established. Similarly, when a PC is playing a music stream, attaching a wire-



Figure 5: Notebook PC configuration: The RFID reader is embedded at the palm rest.



Figure 6: A presenter attaches his/her PDA to the presentation screen to make a connection between them. Then a user can select a file from the PDA display and show it on the presentation screen.

less headphone to the PC would result in wireless music transfer from the PC to the headset. A wireless mouse would also be connected when a user picks up the mouse and places it close to the PC. In both cases, once a wireless connection is established, a user can freely move devices (i.e., the device's position is not restricted to the nearfield channel's range).

4.2 Mobile Presentation

Suppose that you are going to do a presentation in a meeting room where you have never been. Your presentation file is stored on a fileservers at your office. You only bring a mobile wireless device, such as a PDA or a cellular phone. You first put your mobile device on a tag reader installed on the presentation screen. The tag reader then identifies your device and wireless network environment in the room accepts your device and connects it to the network. Then, you select a presentation file on the mobile device's screen, and issue a "show" command. Since the mobile device already knows your target destination is the screen in front of you, your presentation file automatically appears on the screen.



Figure 7: PDA automatically becomes remote commander of target device. (left) An IP address is transmitted from PDA to target device via an infrared beam. (right) The target device responds to the PDA with a description of the device, and user can control it using the PDA. (bottom) The PDA also maintains the access log so user can also connect to the device through the "recently connected" list.

To achieve this scenario, an RFID reader is installed at the side of the presentation screen (Figure 6). When a presenter attaches a wireless PDA (shown in Figure 3) to the screen, an icon appears on the screen to indicate that a connection has been established. Once this operation is done, the presenter can freely walk around without being limited by the RFID reader's sensor range. When a user selected a presentation file on the PDA, a corresponding URL is transmitted to the screen's computer. We also implemented a remote mouse protocol, which is similar to [3], to control presentation screens through the PDA screen. While the Pebbles system requires a user to explicitly enter the target device's address, our system automatically delivers mouse commands to the presentation computer.

4.3 Universal remote commanders

When surrounded by a number of digital devices, control of these devices becomes complicated. Selecting a proper remote commander for each device would be frustrating. Operations that involve two or more target devices, such as showing a movie file stored in a PC on a TV, are difficult to perform because each commander deals only with the corresponding device, and there is no way to carry out inter-device operations. Even the uni-

versal commander, which learns commander sequences and acts like other remote commanders, does not solve this problem because the user still has to select the appropriate mode for each target device.

A remote commander system based on our interaction model addresses this situation and offers a simple operation style. When a user wants to control the device (e.g., a TV set) in front of the user, the user first points at it using his/her mobile device (e.g., a PDA). Then, an infrared beam containing the PDA's IP address is transmitted to the target device, and a wireless connection between the PDA and the target device is established. Next, the target device transmits necessary information for controlling it (such as a command set). Based on this information, PDA becomes a corresponding remote commander for the selected target device. Note that once this connection is set, people can freely move outside of the range limitation of infrared beaming.

Figure 7 shows a typical operation style and a screen shot when a PDA acts as a remote commander. In this case, an IrDA port equipped with the PDA is used to set up a wireless connection. When a connection is established, a web page containing a clickable image map is transmitted back to the PDA (Figure 7 right). As the user clicks on a button on this image, the PDA creates and sends an HTTP request corresponding to the button. The target device (in this case, a PC that functions as a TV) acts as a web server, and handles this request.

When the user wants to show a movie file on the TV screen, the user first selects a movie file on the PDA's screen, and issues a "send" command. While the PDA becomes a TV set's remote commander, it enables the movie to be played on the TV screen. Internally, a handle to the movie file is transmitted to the TV, and the TV retrieves and shows the corresponding movie file from the network server. In our current implementation, a "handle" is simply a URL link, but a handle with authentication information could also be used.

This approach extends the concept of traditional universal commanders in several ways. First, when a connection is established, any prior knowledge about the target device is not necessary. Information about the target device, such as available commands, can be obtained from the target device. Second, since the PDA belongs to each user, a user's personal information can be used while operating the target device. For example, a user can browse his/her movie file list using the PDA's screen, and send its handle to the target device. In this sense, the PDA becomes a "personalized" commander for the target device.

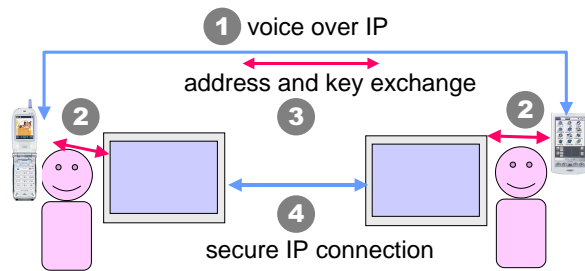


Figure 8: Extensible mobile phones: During a mobile phone conversation (1), user walks up to nearby computer display and makes a connection between the phone and the display (2). The IP addresses and a session key is transmitted over the mobile phone connection (3), displays in front of the phone talkers become a shared screen (4).

4.4 Extensible mobile phones

The proximal interaction model can also be used to connect mobile phones with nearby devices. The sophistication of IP telephony technology advancing quickly, and the industry expects the development of "smart phones" capable of combining the capabilities of mobile computers with telephones, will be released in the near future. These devices will have the ability to exchange short messages, browse web pages, and execute downloaded programs. However, conventional mobile phones are generally isolated from users' nearby computing environment. For example, consider the following situation:

.... Pete is talking to Kane using a cellular phone. They discuss recent changes in design documents. Pete is in his office and looking at a large flat panel display on his desk during the phone conversation. The display shows the design document. Pete wanted to share this information with Kane, but they find it is not a simple operation. Although they are already communicating with IP-based phones, they first have to manually inspect the computers' IP addresses, and exchange them by reading them out, and manually enter these addresses to setup a shared screen application....

The problem with this scenario is that there is no easy way to exchange addresses with nearby devices, even though they are already communicating with IP-based phones. Using our model, this situation could be as follows:

.... Pete is talking to Kane using a cellular phone. When they want to share the display information in front of them, they simply attach their cellular phones to the display sides. Then the RFID readers recognize their cellular phone IDs, and IP communication between the cellular phones and the computer corresponding to the display is established. Then, through the connection between cellular phones, the IP addresses of the display computers are exchanged. The system also exchanges one-time session keys to establish secure communication, and the shared screen session starts without entering any passkeys or addresses...

In this case, two kinds of network connections are established. The first connection is between the user's mobile phone and the nearby display computer. Next, a connection is made between display computers. The second connection is established by first exchanging each devices' address and a session key over the desktop computer (Figure 8), through the created connection (desktop computer A ↔ cellular phone A ↔ cellular phone B ↔ desktop computer B). Since secure communication between cellular phones can be assumed, data transmission over the established phone connection can also be assumed to be secure.

We have implemented this method using PDAs as IP-mobile phones. The PDA with a wireless network card is used as a mobile phone. An RFID tag is attached to the PDA, and an RFID reader is attached to the display side. We use a customized version of IP-phone software that can transmit data packets as well as voice. Microsoft Netmeeting is used as a screen sharing application. Customized gatekeeper software is used to connect both IP-phones and Netmeeting. When a PDA is attached to the display, program on the desktop computer automatically reads the PDA's RFID tag, identifies the PDA's IP address, and transfers this information to the gatekeeper. When a set of this kind of information is transmitted the gatekeeper, it issues a session initiation command to the Netmeeting on both sides with exchanged IP addresses.

5 System Architecture

As shown in the previous examples, we use the combination of physical operations (proximity, pointing, etc.) that initiates wireless connections. These user interface model is achieved using the following elements.

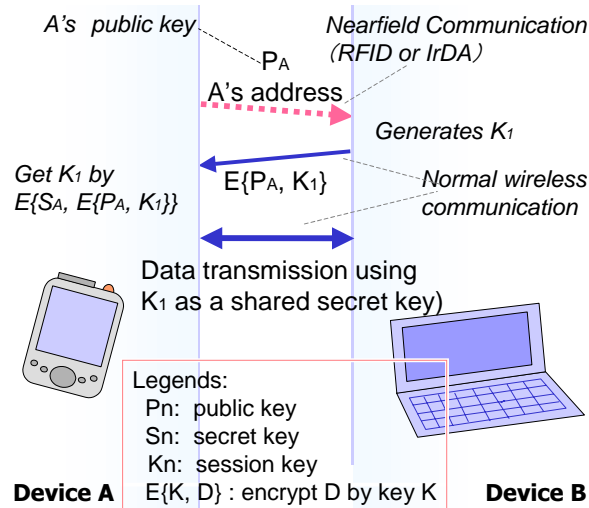


Figure 9: Protocol for establishing secure connection.

5.1 Nearfield communication channels

The nearfield channel is a kind of wireless data transmission method that can be directly controlled by users. "Directly" means that a user can control initiation of data transmission on the nearfield channel with a direct action such as "putting one device within a proximity of another device", or "pointing at one device with the other device." Although a wide variety of range-limited wireless communication technologies might be used, we mainly consider the following types:

RFID: A radio-frequency identification technology, which is widely used in the inventory of ID-cards. As a nearfield channel, one device (normally a mobile device) would have an ID, while the other device would have an ID reader.²

Infrared beaming: Infrared beaming is also widely used in remote commanders or data transmission. The transmission range is from 20cm to several meters. Unlike RFID tags, this transmission is "directional", a user can point at the target device from a distance.

5.2 Standard Channel

The term "standard channel" refers to normal wireless networks, such as IEEE 802.11a, 802.11b, or Bluetooth. There are two types of standard channel communications, ad-hoc mode and infrastructure mode. In ad-hoc mode, two devices establish a network connection after exchanging address information using the nearfield channel. For example, Bluetooth establishes a connection this way. On the other hand, when a device is in an infrastructure mode, such as common in

IEEE 802.11a/b, the wireless connection is already established. In this case, the nearfield channel is used for resolving the address of the devices (i.e., knowing the IP address of the target).

5.3 Communication Protocol

To initiate a network connection, the nearfield channel is used for two purposes: (1) identifying the target device by passing network address such as an IP address, and (2) optionally passing information for a secure connection. Since we consider RFID tags as simplified case of the nearfield channel, the protocol should handle both bi-directional and one-way communications. The actual protocol is shown in Figure 9.

When users initiates a wireless connection, they typically hold the mobile device (i.e., PDA), point it toward the target device (a TV for example). Then, an infrared beam is transmitted from the mobile device to the target device. This beam contains the mobile device's address and a public key corresponding to the mobile device. Upon receiving this information, the target device generates a one-time session key, and completes an encrypted transmission using the received public key. Since only the mobile device can decrypt this key, this reply message can be transmitted through the normal wireless channel. After receiving this message, both sides can communicate with each other using a shared secret session key.

6 Conclusion and Future Research

This paper presents an interaction model for providing an intuitive interface to control and communicate with nearby networked devices. A combination of nearfield communication, such as an RFID or infrared beaming with normal wireless networks offers users a direct way to deal with networked devices, as well as the full capabilities of wireless networking communication. We also present how secure communication can be established using our model, and describe several application examples to show how our model can be used in various situations.

This research is still at an early stage and we need more user feedback. We are currently installing a system in our laboratory environment to study its usability in realistic settings.

In the applications section, we demonstrated how a conventional PDA turns into a remote commander for

² Some RFID technologies provide a bi-directional function that allows communications between two readers, as well as communication between an ID-card and a reader.

the specific target device. In our current implementation, a web page containing a clickable map is transmitted from the target device, and a PDA acts as a web browser. Adding to this simple method, we are also considering dynamically generating commanders GUI from the client side, based on information transmitted from the target device. In this case, the target device transmits the device's description information as a form of XML according to the Universal Plug and Play (UpnP) architecture [6], and the client PDA creates a user interface screen based on this information. Although this method is much more flexible (i.e., the user interface could be tailored depending on the capabilities of the mobile device, such as screen sizes and the number of keypads), automatically generating screen layout is not a simple task.

We also believe that wireless data emission with limited power can also be used as a nearfield channel. For example, a Bluetooth device inquiry message with limited power can be received only by the closest device. Once the target device is identified, normal (full-power) Bluetooth data communication can be established. In this case, the same physical layer (i.e., Bluetooth) is used in both the nearfield and the standard channels. Similarly, sensing the strength of the wireless signal can be used to select the target device.

The other area we are interested in is how the proposed interaction model can manage firewall-based security. Since we regularly carry mobile devices and constantly connecting to surrounding networks, which can be inside or outside of the firewalls, depending on the location of the device and the type of the network. Currently, users have to be responsible for this condition change and must adapt their operation style accordingly. We think our proposed interaction model could simplify this confusing situation.

Acknowledgements

We thank our colleagues at Sony Computer Science Laboratories for the initial exploration of ideas described in this paper. We also would like to thank Toshi Doi and Mario Tokoro for their continuing support of our research.

References

- [1] Dirk Balfanz, D.K. Smetters, Paul Stewart, and H. Chi Wong. Talking to strangers: Authentication in adhoc wireless networks. In *Symposium on Network and Distributed Systems Security (NDSS '02)*, 2002.
- [2] Inc. Bluetooth SIG. Bluetooth sig home page. <http://www.bluetooth.org>.
- [3] Stuart Cheshire. Dynamic configuration of IPv4 link-local addresses. IETF Draft, 2000.
- [4] Infrared data association home page. <http://www.irda.org>.
- [5] T. Kindberg and J. Barton. A web-based nomadic computing system. *Computer Networks*, 35:443–456, 2001.
- [6] Universal plug and play forum. <http://www.upnp.org>.
- [7] Jun Rekimoto. Pick-and-Drop: A Direct Manipulation Technique for Multiple Computer Environments. In *Proceedings of UIST'97*, pages 31–39, October 1997.
- [8] Jun Rekimoto and Yuji Ayatsuka. CyberCode: Designing augmented reality environments with visual tags. In *Designing Augmented Reality Environments (DARE 2000)*, 1999.
- [9] Ben Shneiderman. Direct manipulation: A step beyond programming languages. *IEEE Computer*, 16(8):57–69, 1983.
- [10] Frank Stajano and Ross Anderson. The resurrecting duckling: Security issues for ad-hoc wireless networks. In *7th International Workshop Proceedings on Security Protocols, Lecture Notes in Computer Science*, pages 172–194, 1999.
- [11] Brygg Ullmer, Hiroshi Ishii, and Dylan Glas. mediaBlocks: Physical containers, transports, and controls for online media. In *SIGGRAPH'98 Proceedings*, pages 379–386, 1998.
- [12] R. Want, A. Hopper, V. Falcao, and J. Gibbons. The active badge location system. *ACM Trans. Inf. Syst.*, January 1992.
- [13] Roy Want, Kenneth P. Fishkin, Anuj Gujar, and Beverly L. Harrison. Bridging physical and virtual worlds with electronic tags. In *CHI'99 Proceedings*, pages 370–377, 1999.
- [14] Mark Weiser. The computer for the twenty-first century. *Scientific American*, pages 94–104, September 1991.
- [15] Zero Configuration Networking. <http://www.zeroconf.org>.