The IR Ring: Authenticating Users' Touches on a Multi-Touch Display

Volker Roth, *Philipp Schmidt*, *Benjamin Güldenring*

Secure Identity Research Group
Freie Universität Berlin
Takustr. 9, 14195 Berlin, Germany
{volker.roth, philipp.schmidt, benjamin.gueldenring}@fu-berlin.de

ABSTRACT

Multi-touch displays are particularly attractive for collaborative work because multiple users can interact with applications simultaneously. However, unfettered access can lead to loss of data confidentiality and integrity. For example, one user can open or alter files of a second user, or impersonate the second user, while the second user is absent or not looking. Towards preventing these attacks, we explore means to associate the touches of a user with the user's identity in a fashion that is cryptographically sound as well as easy to use. We describe our current solution, which relies on a ring-like device that transmits a continuous pseudorandom bit sequence in the form of infrared light pulses. The multi-touch display receives and localizes the sequence, and verifies its authenticity. Each sequence is bound to a particular user, and all touches in the direct vicinity of the location of the sequence on the display are associated with that user.

Keywords: Authentication, multi-touch.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Input devices and strategies

General terms: Design, Security, Human Factors

INTRODUCTION

Multi-touch displays are enjoying an increasing popularity, as is evident by the many reports of private and commercial multi-touch projects on the Web. Over time, the technology will likely be commodified and deployed as a common tool for multi-user interaction, where its advantages come to bear best. In the wake of that success, pranksters and miscreants will certainly follow, trying to adapt their schemes of fraud and malice to the new technology and its uses. This in turn gives rise to the question how *access control* and *authentication* will be handled on a multi-touch display (MTD).

For example, a user Eve shall not be able to delete the data of another user Alice with whom she shares access to a MTD.

UIST'10, October 3-6, 2010, New York City, NY, USA.

Figure 1: Shows the authenticated touch area underneath a user's hand. The IR Ring which identifies the user and authenticates the location is worn underneath the hand (not visible).

Even more importantly, Eve shall not be able to send messages to third parties in the name of Alice while Alice is not looking because this action cannot be remedied easily with an *undo* operation. In addition to adverse scenarios, the authentication of users' touches would also be useful in multi-user games, especially competitive ones, and as a means to personalize multi-touch applications.

Towards a practical exploration of these questions, we built a small and lightweight low-powered device that a user can wear like a ring. The device *uniquely and simultaneously identifies itself and its location* to the MTD by emitting a cryptographically generated pseudorandom signal in the infrared spectrum. The MTD locates the signal on its surface in the same fashion it locates touches, and it associates all touches in the direct vicinity of the authenticated location with the user who is associated with the device (see Fig. 1). The probability that an adversary forges a valid signal decreases exponentially in the signal length. In contrast to other "all or nothing" security mechanisms, our approach supports trade-offs between computational resources, security and usability. As an illustration, we describe a security policy that attaches variable authentication probabilities to user interface elements with different sensitivity levels.

RELATED WORK

The pointing end of the *XWand* [11] 3D pointing device features a IR light that its tracked and triangulated by means of

two IR cameras located in a room. This anchors the local coordinate system of the XWand within the room's coordinate system. By pointing the XWand at other devices located at well-specified positions within that room, a user can control said devices e.g., the user can remotely dim room lights.

The *SurfaceFusion* project by Olwal et al. [7] combines computer vision based tracking with RFID based identification. When an object is placed onto the MTD, the RFID component identifies the object, and a computer vision component locates and tracks its position on the table. Olwal et al. did not use RFID for precision object tracking because of the technical complexity of such an approach. For example, RFID measurements suffer from interferences among multiple tags and the necessary equipment is not available off-the-shelf. Our approach, on the other hand, has a low complexity and is precise and reliable. Schöning et al. [8] describe an approach whereby a mobile phone authenticates its position on a back-projection display by simultaneously flashing a light and by transmitting a digitally signed time-stamp.

Both approaches mentioned above rely on the *synchronicity* of two events occurring over two separate channels, which means that only one object can be handled at a time. For illustration, let (a, l) be an authentication and location event pair. Then the simultaneous occurrence of (a, l) and (a', l') is indistinguishable from the simultaneous occurrence of (a', l) and (a, l'), which means that a device can be authenticated but not its position. Our approach, on the other hand, solves this problem by merging the channels for authentication and location, rendering them inseparable.

The project that is closest to ours is *DiamondTouch* by Dietz et al. [4]. The DiamondTouch table uses capacitive coupling between a chair and the table to distinguish the touches of different users. Authentication can certainly be added to the capacitive channel. On the other hand, our mechanism allows users to move around the table freely, and it is compatible with popular multi-touch technology. A fingerprint scanning display would be another interesting alternative. Such a capability would also enable more precise touches than are typical at the time of writing [5]. Sugiura et al. [10] already explored interaction techniques that could be implemented with such a device. However, attractive as the idea is, fingerprints are easily copied and, without livelihood detection, the security would be limited.

Light has been explored as a location-limited communication channel for cryptographic transmissions e.g., in [2], but tracking has not played a role in this work. Corner [3] also leveraged location-limited channels for transient user authentication by verifying the presence of a user's personal mobile device (the IR Ring would be well-suited for that as well).

THREAT MODEL AND SECURITY POLICIES

We assume that multiple users can share the display while accessing private resources such as electronic mail or social networking accounts. The threat we address is that of a user who interacts clandestinely with user interface controls of another user's applications by touching them. Since some interactions are desirable while others are clearly not, we assume that we can formulate a security policy that maps allowed actions to the user interface components that trigger these actions.

For clarity of illustration, we define a simple policy by means of an access control matrix M where the rows represent users S and the columns represent user interface controls O. Each entry in the matrix $M(s, o)$ specifies in a boolean fashion whether user s is allowed to activate the control o. This would be a typical approach, and it requires that all users are authenticated at all times at the security level of the most sensitive control. One example of this model would be the mechanism Schöning et al. [8] implemented on a mobile phone. The phone flashes a light to specify its position and at the same time computes and transmits a digital signature via Bluetooth. The security of this approach is maximal starting from the first authentication but has comparatively high costs for continuous operation.

In our model, we extend the matrix to rational entries in the range $[0, 1]$. A value of $p = M(s, o)$ means that a touch on control o is allowed if and only if the touch can be attributed to user s with a probability greater than p. Our enforcement mechanism, of which the IR Ring is a component, expends comparatively few resources to transmit a continuous stream of pseudorandom bits at a low rate by means of a blinking IR LED. The level of security starts low but compounds over time at an exponential rate. This model enables us to achieve very high levels of security with a very small and very resource-limited device. The disadvantage is that sensitive controls may not be activated until the device is "visible" to the MTD sufficiently long to exceed the security threshold set forth by the policy.

A possible threat that we do not currently address is that of an adversary pointing a laser from remote at a position within the authenticated touch area of another user, which may trigger a touch event on a computer vision based MTD. This threat must be countered at the environmental level, if necessary. An adversary may also attempt to capture reflections of an IR Ring's light pulses with the goal of relaying them to the MTD, all while the user of the IR Ring is elsewhere in the room and away from the table. This threat can be countered e.g., by generating IR "noise" in the room. Another alternative would be to use additional sensors to detect when the ring is away from the MTD and to suspend the ring's LED in these cases. In a similar fashion, the IR Ring can be invalidated when taken off the finger.

Lastly, the LED of an IR Ring generates blobs that are clearly discernible in the IR images, which limits the opportunity for cross-talk between two IR Rings as a source of errors.

HARDWARE

The IR Ring features a TI CC430F6137 microcontroller with integrated AES hardware encryption and a RF transceiver, mounted on a 30 mm $\times$ 24 mm printed circuit board (PCB) together with one IR, red, green, and blue LED, an accelerometer, a light sensor, and a push button (see Fig. 2). A battery attaches to the bottom of the circuit board and lasts for an estimated 8 hours of continuous operation. The PCB can be made even smaller by omitting the JTAG and serial interfaces that we currently use for ease of debugging.

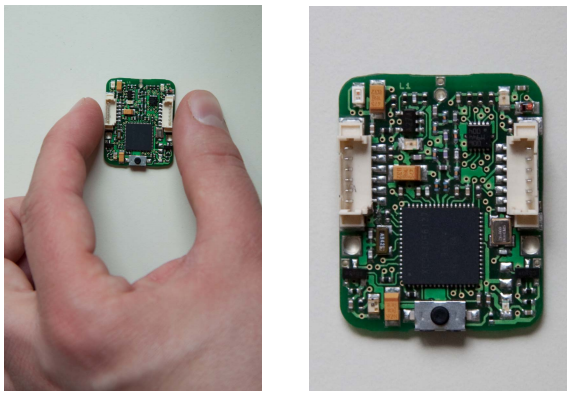

Figure 2: Shows the IR Ring prototype circuit board.

INFRARED TRANSMISSION

The IR Ring transmits a Manchester-encoded (see e.g., [1]) pseudorandom bit sequence by means of *on*/*off* states of its IR LED. The Nyquist-Shannon sampling theorem [9] bounds our maximum data rate from above at half the receiver's sampling frequency. Since the Manchester encoding requires the transmission of two bits (i.e., one *cycle*) to encode one data bit, this further reduces our theoretical maximum rate to a fourth of the receiver's sampling frequency. For example, with a 60 Hz camera setup we cannot expect a data rate better than 15 bits. In practice, the rate is typically lower than the theoretical maximum. One advantage of the Manchester encoding is that the receiver can recover the sender's clock, and use this information to prevent bit shift errors due to lost bits. This significantly simplifies the matching of the transmitted sequence against the sequences the receiver expects. Another important property of the Manchester encoding for our application is that a zero bit must occur after at most two consecutive 1 bits (and vice versa). This allows us to distinguish a touch from the IR LED based on time: a blob that persists for at least one and a half cycles must be a touch. Shorter touches register as spurious IR Ring events and can be eliminated quickly by setting a minimal security threshold. At the same time, the encoding guarantees that the IR LED can be tracked at least once every one and a half cycles.

DECODING AND MATCHING ALGORITHMS

The first stage in our decoding and matching pipeline is a software that extracts blobs from a video stream (see Fig. 3 for illustration). We have successfully used the *Community Core Vision* framework in conjunction with a PS3 Eye camera, and, in another setup, the *Wii Remote* for this purpose. In both cases, the blob detection is stateful i.e., the output is a unique blob ID id, a point p and a timestamp t.

This information is input to a *switch*. The switch first determines whether the blob is likely a touch or a part of an authentication sequence (briefly, a *blip*). The blob is classified as a touch and forwarded to the TUIO event emitter if the ID is already known and the time since the switch last saw this ID is greater or equal to one and a half cycles of the Manchester decoder. Otherwise, if the ID is known, the switch classifies the blob as a blip and forwards it to the *decoder* for further processing.

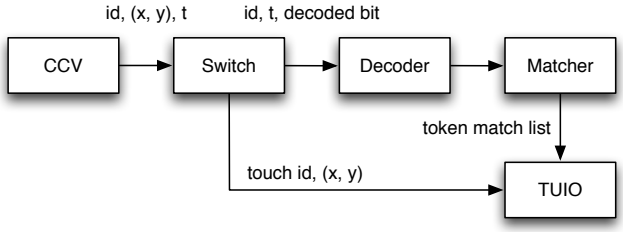

Figure 3: Illustrates the our software architecture.

If the ID is not yet known then the blob could be a new touch, or it could be the blip of an ongoing or a new authentication sequence. The switch finds the ID id' of the closest known point p' for which the Euclidean distance between p and p' is less than a given threshold. If such an id' exists then the switch updates all values of id' to id and forwards the blip to the decoder. Otherwise, the switch creates a new entry for the blob with the given ID id and returns.

The *decoder* attempts to extract a data bit based on the Manchester decoding algorithm. The output can be a data bit, a wildcard (a data bit for which the value cannot be decided), or a symbol that signals loss of synchronization. The decoder passes the output to the matcher.

The *matcher* maintains a current window of the sequences of all registered IR Rings. The window is extended on demand (i.e., if all bits are matched) and based on time (i.e., if there is a synchronization loss). For each ID and sequence pair, the matcher performs the following steps. If the output is a data bit or a wildcard then the decoder appends it to the decoded sequence, matches the decoded sequence against the expected sequence for all *valid shifts*, and calculates a score for each shift. The valid shifts are initially all shifts within a maximum shift window e.g., eight bits into the past or future. As soon as the score of one valid shift exceeds a given threshold then all other shifts are removed. A valid shift is also removed if there is a mismatch i.e., the exclusive-or of the received bit and the expected bit is one. Otherwise, a match occurred and the score is updated. The score is calculated as $1 - 2^{m-n}$ where n is the length of the received sequence and m is the number of wildcards in the received sequence. Less formally, we treat wildcards as potential matches but they do not lead to an increase of the score.

The matcher outputs a list of quadruples of the form

$$(id_0, p_0, \text{ring}_0, \text{score}_0), \ldots, (id_k, p_k, \text{ring}_k, \text{score}_k)$$

for $0 \leq i \leq k$ where id_i is a unique ID of the authenticated position p_i, ring_i is the unique ID of an IR Ring, and score_i is the probability that position p_i with ID id_i represents the IR Ring with ID ring_i. Finally, the *TUIO emitter* takes the touch and authentication records and emits TUIO protocol messages for them (we have extended the TUIO 1.1 protocol to support authentication events). A garbage collection process clears all records that become inactive e.g., because no blip was recorded within one and a half Manchester decoder cycles. The TUIO events must be interpreted by the application or system that intends to enforce multi-touch authentication. In order to prevent tampering with the events during

transmission, a secure channel must be established between the TUIO emitter and the application or system using e.g., OpenSSL. We have not built such support yet.

CRYPTOGRAPHIC SECURITY

Our current firmware implementation is limited to cycling through a fixed pseudorandom bit sequence. Our next step is to generate the sequence by encrypting a stream of zeroes using the AES cipher in randomized counter mode with a random initialization vector. The key will be derived from a master key that is exchanged with the MTD system through a trusted channel. We omit proofs of security because they are outside the scope of this paper and the involved mechanisms are well understood [6]. The important detail is that no polynomially bounded adversary (a standard assumption) given n bits of the output can predict the $n+1$'th output bit with a probability non-neglibly better than 1/2. The probability of one IR Ring being mistaken for another is governed by the *Birthday Paradox*, which means that, given a constant number of rings, the probability is bounded from above by a function that decreases exponentially in the length of the sequence.

TEST ENVIRONMENT AND INITIAL TESTS

We built a "poor man's multi-touch table" to perform the initial testing of our IR Ring prototype. The table consists of a 45cm $\times$ 35cm PMMA surface on top of a cardboard box. The surface is lit with 24 IR LEDs that are glued to its sides (we used 880nm SFH421 SMD LEDs powered with 60 mA). We did not filter out IR light from the projection. A PS3-Eye camera with its IR filter replaced by a IR pass filter served as the sensor for the table and for the IR Ring. We set the camera to a 320 $\times$ 240 pixel resolution at 60 Hz and used a patched version of the *Community Core Vision* framework as our image processor. The software ran on a Mac Mini with a 2.0GHz Core 2 Duo Processor and 4GB RAM. Due to camera driver issues we were not able to leverage the maximum frame rate of 120 Hz supported by the PS3 Eye. Before we built the table prototype, we tested the IR Ring using a Wii Remote via Bluetooth as the sensor. Although the Wii Remote supports a frame rate of 100 Hz, the extensive jitter we experienced from the Bluetooth connection effectively reduced the usable data rate of our IR Ring to the same rate we achieved with the PS3 Eye at 60 Hz. For this reason, we continued our testing with the prototype table.

At a data rate of 5 bits per second we measured 10% wildcards, 0.6% bit flips and 1.7% spuriously inserted bits, all of which occurred as a consequence of consecutive wildcards. Further investigation suggests that this problem can be eliminated by means of small software changes. On average, it took 2.4 seconds to achieve an authentication probability of 0.997 (8 data bits) with a minimum of 1.8 seconds and a maximum of 4.8 seconds ($s^2 = 0.28$ seconds).

DISCUSSION AND CONCLUSIONS

We presented the IR Ring, a cryptographic ring-like device (reminiscent of the infamous Java Ring by Dallas Semiconductors) suitable to authenticate the touches of a user on a multi-touch display in a fashion that affords a trade-off between computational power, security and usability.

For tracking purposes, the IR Ring's LED must point towards the table, which may lead to occlusion e.g., if the user curls his fingers for convenience and to avoid hand fatigue. This problem can be mitigated by wearing the IR Ring on the thumb or by tracking it from above. Tracked touches can remain at the authenticity level they had before the occlusion.

Our test setup included a cheaply built "poor man's FTIR table." Given the ad hoc nature of our table prototype we are satisfied with the initial outcome. Furthermore, our current results likely represent a lower bound of what is achievable. However, the overall performance can be improved in several ways. For example, we can eliminate the touch latency imposed by our decoding mechanism by using different wavelengths for the IR Ring and for the touch signals, and by using two cameras with appropriate band pass filters (one camera for each wavelength).

Acknowledgments We would like to thank the following: Konstantin Käfer and his team at HPI at Potsdam University for providing their expertise, Texas Instruments for providing parts, Achim Liers assisting with the system design, and the reviewers for their insightful comments.

REFERENCES

1. A. Ageev. Manchester decoder and clock recovery module for FPGA prototype of active RFID tag. Technical report, Berkeley Wireless Research Center, UC Berkeley, Dec. 2009.
2. D. Balfanz, D. K. Smetters, P. Stewart, and H. C. Wong. Talking to strangers: Authentication in ad-hoc wireless networks. In *NDSS*, 2002.
3. M. D. Corner. *Transient Authentication for Mobile Devices*. Ph.d. thesis, University of Michigan, Aug. 2003.
4. P. H. Dietz and D. Leigh. DiamondTouch: a multi-user touch technology. In *UIST*, pages 219–226, 2001.
5. C. Holz and P. Baudisch. The generalized perceived input point model and how to double touch accuracy by extracting fingerprints. In *CHI*. ACM, 2010.
6. J. Katz and Y. Lindell. *Introduction to Modern Cryptography*. Chapman & Hall/CRC, 2008.
7. A. Olwal and A. D. Wilson. SurfaceFusion: unobtrusive tracking of everyday objects in tangible user interfaces. In *Graphics Interface*, pages 235–242, 2008.
8. J. Schöning, M. Rohs, and A. Krüger. Using mobile phones to spontaneously authenticate and interact with multi-touch surfaces. In *Proc. Workshop on Designing Multi-Touch Interaction Techniques for Coupled Public and Private Displays*, Naples, Italy, May 2008.
9. C. E. Shannon. Communication in the presence of noise. *Proc. IRE*, 37(1):10–21, Jan. 1949.
10. A. Sugiura and Y. Koseki. A user interface using fingerprint recognition: holding commands and data objects on fingers. In *UIST*, pages 71–79. ACM, 1998.
11. A. Wilson and S. Shafer. XWand: UI for intelligent spaces. In *CHI*, pages 545–552. ACM, 2003.