

# Smart Signage: A Draggable Cyber-Physical Broadcast/Multicast Media System

JAMES SHE<sup>1,2</sup>, JON CROWCROFT<sup>1</sup>, HAO FU<sup>2</sup>, AND PIN-HAN HO<sup>3</sup>

<sup>1</sup>Computer Laboratory, University of Cambridge, Cambridge CB2 1TN, U.K.

<sup>2</sup>Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong

<sup>3</sup>Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

CORRESPONDING AUTHOR: J. SHE (eejames@ust.hk)

**ABSTRACT** Digital displays, as the replacement of traditional static signs, have gained increasing popularity for out-of-home advertising. Latest advancements in smartphone, wireless communication, and digital display technologies make it possible to design new interactive signage systems linking the digital content with the physical digital displays. Although recent studies have demonstrated the trend of cyber-physical interaction, they are not generally scalable for multiple users, and none of them support interaction with multiple displays in one location. Smart Signage, a “draggable” cyber-physical broadcast/multicast (B/M) media system is proposed in this paper. With a novel cyber-physical B/M protocol that synchronizes the content on the digital displays with the smartphones, it supports one-to-many interaction by allowing multiple users acquiring content from one display with a “dragging” hand gesture. With the embedded display orientation information in the same protocol, Smart Signage supports many-to-many interaction by allowing users obtaining content from the digital display they are pointing at with their smartphones. Users’ quality of experience, which is characterized by the response time, is carefully studied in this paper to guarantee the performance of this cyber-physical interactive display system.

**INDEX TERMS** Interactive display, cyber-physical system, draggable broadcast/multicast media, smartphones, advertising.

## I. INTRODUCTION

Traditional static signs have been used as a convenient way to broadcast advertisements to a large number of targeted audiences who pass by the signages located at public/semi-public areas. As one of the examples, tear-off advertisements are commonly used with some creative design which is shown in Fig. 1 [1]. A tear-off advertisement allows individuals to physically tear off a part of the signage to keep some specific information (e.g., contact number) of the advertisement for follow-up actions.

Digital displays, as the replacement of traditional static signs, have gained increasing popularity, especially for out-of-home advertising. Although easily maintained and updated, digital signages generally do not provide interactive services to the audiences. The pervasiveness of the smartphones makes it possible to implement new interactive digital signage systems for advertising purpose. These interactive display systems, as shown in Fig. 2, can generally be abstracted by three key components, which are *display*,



**FIGURE 1. Examples of tear-off advertisement: (a) conventional tear-off advertisement; (b) tear-off advertisement with creative design [1].**

*system* and *interactive modality*. The *display* represents a visualization equipment, which could be a flat-panel LCD unit, a screen with projector, a digital billboard, a LED matrix display, a 3D display, a PC monitor, or any emerging flexible and transparent display. The *system* is an embedded media

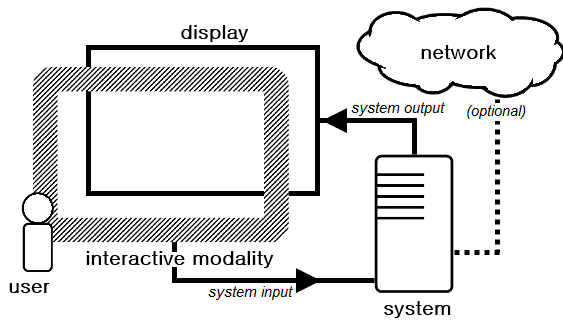


FIGURE 2. General interactive display system architecture.

playback and control system with/without network connectivity that coordinates the display and interactive modality. The *interactive modality* is the technical method, physical scale and social style such that users can interact with the displayed content.



FIGURE 3. Examples using: (a) QR code [9]; (b) U-tie [10].

Although the digital displays today can present various media contents to the audiences, most of their interaction modalities so far are still limited to one-to-one interaction. Bluetooth [2]–[4] is a popular technology for interactive modality, but the master-slave structure of the piconet only allows up to 7 users, which is a great limitation on the scalability. Cameras have been used as interactive modality with image processing, where the cameras can be system-side cameras [5], [6] or device-side cameras [7], [8]. As a special case of cameras, QR code is widely adopted in recent years. To create a distance that allows a large number of users scanning the QR code concurrently, a large and unappealing QR code is needed, which visually compromises the limited advertising area as shown in Fig. 3(a) [9]. With powerful capabilities of smartphones today, novel image processing techniques are employed in commercial applications like U-tie, as shown in Fig. 3(b) [10]. It provides multiple user interaction and information retrieval through the images captured by the users' smartphones. Unfortunately, this approach requires Internet access and relatively long processing time. Near field communication (NFC) [11], [12] is an emerging technology to provide one-to-one interaction that allows users to quickly collect information, but it requires the smartphones to be within a few centimeters' distance. Wi-Fi is another popular technology for interactive modality. Compared with previous technologies, it has the advantages of high throughput

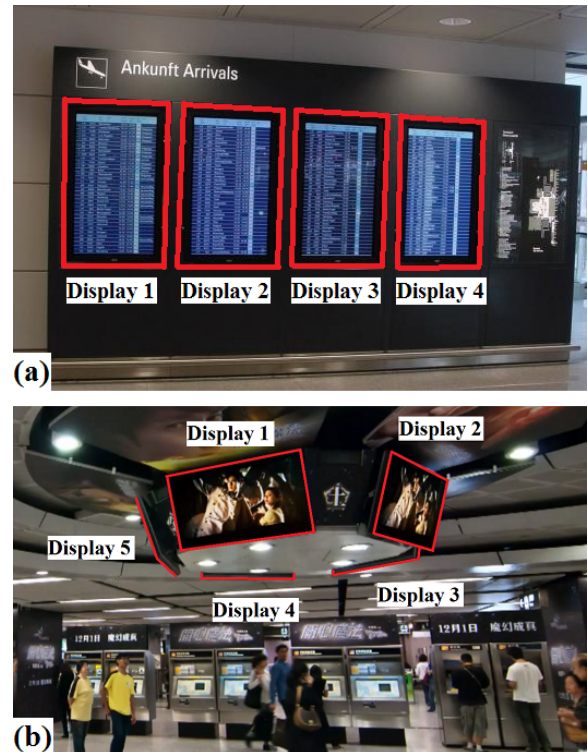


FIGURE 4. Multiple digital displays in one location: (a) airport information displays; (b) advertising campaign.

and flexible interaction range. Some of the examples using Wi-Fi [13]–[15] have demonstrated the possibility of using smartphone accelerometers to recognize users' hand gestures as intuitive interactive modalities.

With the increasing number of deployed digital signages, it is possible that there are multiple signage in one location. Fig. 4(a) shows the information displays at an airport, where four displays are placed in parallel on a wall. Fig. 4(b) shows the setup of displays at a subway station for an advertising campaign, where five digital displays were used. However, none of the previous mentioned systems addressed the problem of interaction between smartphones and multiple digital displays, which calls for new pervasive computing techniques, which should be sensor-based, multimodal and touchless [16]. The fact that digital signages require the users to be within physical proximity to interact with digital information makes the technique cyber-physical in nature.

Motivated by the above observations, Smart Signage, a “draggable” cyber-physical broadcast/multicast (B/M) media system is proposed in this paper. With a novel cyber-physical B/M protocol that synchronizes the content on the digital displays with the smartphones, it supports one-to-many interaction by allowing multiple users acquiring content from a display with “dragging” hand gesture. With the embedded display orientation information in the same protocol, Smart Signage supports many-to-many interaction by allowing users obtaining content from the digital display

**TABLE 1. Advantages of draggable cyber-physical B/M media.**

Technologies \ Metrics	Bluetooth	System-side camera	QR code	Device-side camera	NFC	Smart Signage
Simultaneous user interaction	×	×	×	×	×	✓
More flexible distance	×	×	✓	✓	×	✓
More intuitive interaction	×	✓	×	×	✓	✓
Less visual compromise	✓	×	×	✓	✓	✓
Higher throughput	×	×	×	×	×	✓
Mobility support	×	×	×	×	×	✓

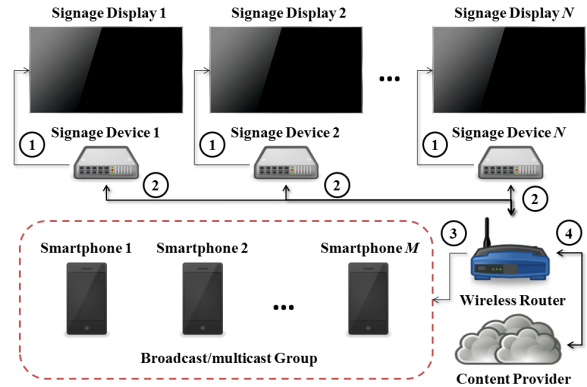
they are pointing at with their smartphones. Users' quality of experience (QoE) which is characterized by the response time, is studied in this paper. As a cyber-physical system, Smart Signage is designed in a way that the users are able to interact with digital content of good quality while experience almost no delay during the interaction process even in a location with multiple displays.

To the best of our knowledge, Smart Signage has advantages over other existing approaches in terms of the scalability to engage a large number of users, more flexible interaction range, more intuitive interaction, less visual compromise on the presentation of the advertisement, higher throughput, supporting interaction with mobile users, and supporting interaction with multiple displays. These are important improvements compared with conventional interactive modalities used in interactive display systems such as Bluetooth, cameras with image processing, QR code, and NFC, as summarized in Table 1.

The subsequent paper is organized as follows. Section II introduces the proposed Smart Signage system; Section III gives the performance analysis of the system; Section IV shows the implementation and experiment results; Section V concludes this paper.

**II. THE PROPOSED SMART SIGNAGE SYSTEM**

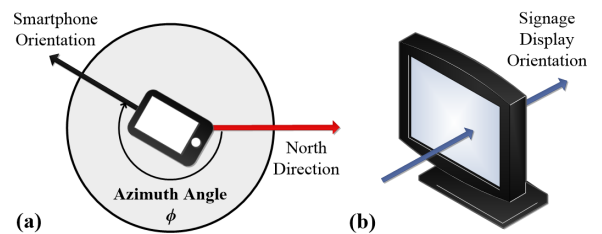
Fig. 5 shows the system architecture of the Smart Signage system. It consists of  $N$  signage displays, each connected to a signage device, a wireless router, a content provider and  $M$  smartphones. There are four types of links illustrated in the figure. The signage device controls the content displayed on the signage display through a video output link ①. The signage devices transmit B/M data packets to the wireless router using wireless data links ②. The wireless router relays these packets through a wireless data link ③ to the smartphones that have joined a particular B/M group, which is the same as the one the signage devices have joined. The wireless router also is responsible for fetching the content data and control sequence for the signage devices from a content provider through an Internet data link ④ and forwards the information to the signage devices through wireless data links ②. Although the current system design uses Wi-Fi in the implementation, any wireless standard supporting B/M radio signals is also applicable. The signage device downloads updated content for the display from a content provider through the Internet access.



**FIGURE 5. System architecture of Smart Signage.**

**A. SMARTPHONE ORIENTATION AND SIGNAGE DISPLAY ORIENTATION**

In order to support the interaction between smartphones and multiple signage displays, a sensor-based technique is used to differentiate signage displays. As location information is not easy to obtain especially for indoor scenarios, this approach only utilizes the orientation information of the smartphones and signage displays.



**FIGURE 6. Illustration of (a) smartphone orientation and (b) signage display orientation.**

Many commercial smartphones nowadays have embedded orientation sensors, which measure the postures of the smartphones in 3D space. An orientation sensor can be constructed from an accelerometer and a magnetic sensor [17]. The data measured by the accelerometer and magnetic sensor are processed to get the azimuth angle and inclination angle of the smartphone. In this paper, only the azimuth angle is utilized to determine the orientation of the smartphone, which is illustrated in Fig. 6(a). In this figure, the disk represents the plane which is perpendicular to the direction of Earth's gravity. The north direction is the direction of the Earth's magnetic north projected on the plane. The *smartphone*

*orientation* is the direction of the smartphone projected on the same plane. The smartphone orientation is quantified by the azimuth angle  $\phi$  measured clockwise from the north direction to the smartphone orientation. The azimuth angle used to measure the orientation of the user's smartphone  $\phi$  varies within  $[0, 2\pi)$ .

The *signage display orientation*, as shown in Fig. 6(b), is defined as the direction that is perpendicular to the signage display, and is pointing into the display surface. When a user is standing in front of the signage display and pointing his/her smartphone at the signage display, the smartphone orientation aligns with the signage display orientation, which means the signage display orientation can be quantified by the same azimuth angle measured by the smartphone. The azimuth angle used to measure the orientation of the signage display is denoted as  $\theta$ . By assuming there are multiple signage displays in one location with different orientations, they can be differentiated by an array of different azimuth angles,  $\{\theta_n | n = 1, 2, \dots, N\}$ , each of which representing a particular sign  $S_n$ , where  $n = 1, 2, \dots, N$ .

When a user is walking around in the location with multiple digital displays, he/she does not always stand in front of the signage display, which the user is intended to interact with. So the smartphone orientation is not always the same with the signage display orientation when the user is pointing the smartphone at the signage display. To resolve this issue, each sign  $S_n$  is assigned with an azimuth angle range,

$$R_n = \begin{cases} [\omega_{n-1}, \omega_n) & \text{for } n \geq 2 \\ [\omega_N, \omega_1) & \text{for } n = 1, \end{cases} \quad (1)$$

where,

$$\omega_n = \begin{cases} \frac{1}{2}(\theta_n + \theta_{n+1}) & \text{for } n < N \\ \frac{1}{2}(\theta_N + \theta_1 + 2\pi) - 2\pi \lfloor \frac{(\theta_N + \theta_1 + 2\pi)}{4\pi} \rfloor & \text{for } n = N. \end{cases} \quad (2)$$

In other words, the available azimuth angle space between two adjacent displays are equally divided and assigned to each of the two displays. With the adoption of azimuth angle ranges, as long as the smartphone orientation is within the azimuth angle range of the intended sign, the smartphone will correctly interact with the one which the user is pointing at. As an illustration of the approach differentiating signages, Fig. 7 shows an example where three signage display are deployed and the user is trying to interact with one of the signages using smartphone. It is easy to observe that the smartphone orientation  $\phi$  is not the same as the signage display orientation  $\theta_2$  when the user is pointing the smartphone at the signage display of  $S_2$ . However, as  $\phi$  is within the azimuth range, namely  $[\omega_1, \omega_2)$ , assigned to  $S_2$ , the smartphone still can correctly presume the user is interacting with  $S_2$ .

### B. A CYBER-PHYSICAL BROADCAST/MULTICAST PROTOCOL

In order to support one-to-many interaction by allowing multiple users acquiring content from one display concurrently, one of the design objectives of the system is to synchronize

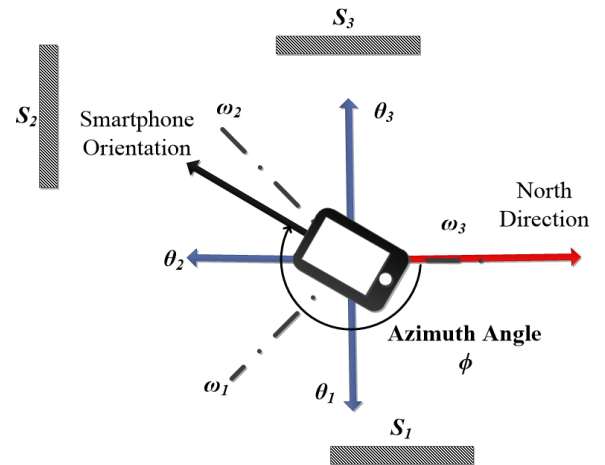


FIGURE 7. Illustration of the azimuth angle ranges.

the content on the digital displays with the smartphones. To achieve this, a one-directional B/M protocol is proposed. Compared with some traditional B/M protocols such as [18], the protocol of the Smart Signage system has several special mechanisms to guarantee user experience. Since the event of “dragging” hand gesture at the smartphone side is equivalent to a request for a content on the display (or a “draggable” file from the signage device), which is difficult to predict the timing, it is preferable to buffer all the “draggable” files on the smartphones. The protocol also embeds the information about what is displayed on the signage display so that the smartphones are fully aware.

In order to support many-to-many interaction the cyber-physical B/M protocol is designed to incorporate the orientation information of the display. The data packet is shown in Fig. 8, in which an *Orientation* header field is added. For a B/M packet sent from sign  $S_n$ , this orientation header field is an integer indicating the azimuth angle  $\theta_n$  that represents the signage display orientation, where  $n = 1, 2, \dots, N$ . Utilizing the orientations  $\{\theta_n | n = 1, 2, \dots, N\}$  instead of the azimuth angle range boundaries  $\{\omega_n | n = 1, 2, \dots, N\}$  makes the system more adaptive. When new displays are added to a specific location, new boundaries of the azimuth angle ranges can be easily computed by the smartphones based on the received signage display orientation information, which requires no change to the packets' orientation header fields of all the displays.

Signage ID (4B)	Orientation (4B)	Showing ID (4B)	Sending ID (4B)	Data Header (4B+4B+4B)
Data Payload (Maximum 1024B)				

FIGURE 8. Data packet structure.

For the rest of the fields in the data packet, *Signage ID* is a unique identification of the signage device; *Showing ID* is the identification of the content being displayed on the signage display; *Sending ID* is the identification of the content being



carried in this data packet; *Data Header* includes the size of the content with the specified Sending ID, the total number of packets of the sending content and the sequence number of the data fragment of this content encapsulated in this packet; *Data payload* contains the data fragment limited to the size of 1024 bytes.

### C. SOFTWARE DESIGN OF SIGNAGE DEVICE AND SMARTPHONE

*Signage device:* As shown in Fig. 9, once a signage device starts running, it first connects to the wireless access point and then the program splits into two subroutines. Subroutine 1 accesses the Internet and constantly checks for updates for the content changes pushed by the content provider. If there is an update, the signage device will download the new content and corresponding control sequence to the local storage. Subroutine 2 joins a particular B/M group and continuously sends the packets to this group. The packets have the same format defined by the orientation-enabled cyber-physical B/M protocol as illustrated in Fig. 8. Subroutine 2 also updates the content showing on the signage display according to a play list generated from the control sequence downloaded from the content provider.

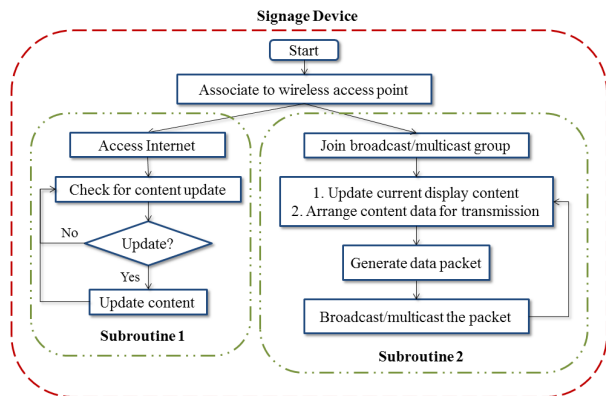


FIGURE 9. Software design of the signage device.

*Smartphone:* As shown in Fig. 10, once the application on smartphone starts running, it first connects to the wireless access point and joins the same B/M group as the signage devices, and then the program splits into two subroutines. Subroutine 1 continuously listens to the packets in the B/M group, and extracts header fields of Signage ID, Orientation, and Showing ID, which are used to form a database that keeps track of the information and status of all the displays at this location. Once the database is available, the application calculates the azimuth angle range for each signage display for later use. Subroutine 1 also processes the data payloads of the packets and stores them as content files in local buffer storage. Subroutine 2 uses the accelerometer data to detect a “dragging” hand gesture, which is treated as the signal issued by the user to interact with the digital sign. If the hand gesture is successfully detected, the application checks with the orientation sensor data and the calculated azimuth angle

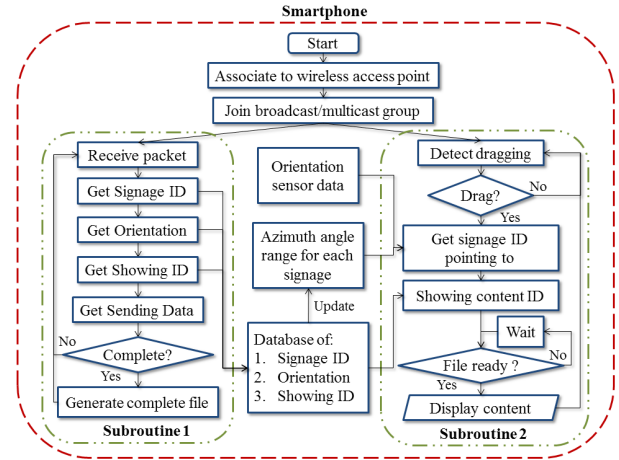


FIGURE 10. Software design of the smartphone.

ranges to determine which sign the smartphone is pointing at. After getting the Signage ID, the application obtains the Showing ID of the same sign, and checks whether the content with the specified identification has been buffered in the local storage. If the content file is ready, the content will pop up on the smartphone’s display to complete the interaction process.

### D. DETECTION OF “DRAGGING” HAND GESTURE

Note that the smartphone only detects one simple dragging gesture as shown in Fig. 11. Although conventional hand gesture recognition generally requires gesture spotting and gesture segmentation [19], the fact that only one simple gesture is used in the proposed system makes it possible to detect the dragging event by simply checking the range of the values of tri-axis accelerometer when total acceleration,  $\bar{A} = \sqrt{A_x^2 + A_y^2 + A_z^2}$  (where  $A_x$ ,  $A_y$  and  $A_z$  are the acceleration values of X, Y and Z-axis of the accelerometer respectively), reaches the maximum. The measurements of the tri-axis accelerometer when the dragging gesture is performed is shown in Fig. 12. The data is combined from 20 users who were instructed to perform the dragging gesture repeatedly. It is easy to observe that  $\bar{A}$  reaches maximum at the 15th data point. Hence the corresponding range of values of each axis at the maximum total acceleration, namely  $R_{AX}$ ,  $R_{AY}$  and  $R_{AZ}$ , can be determined from the figure by covering the most representative 80% of the entire data. The most

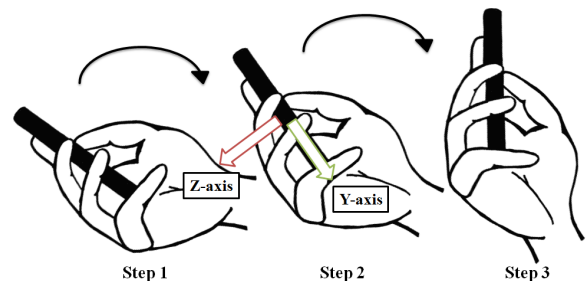


FIGURE 11. Illustration of the “dragging” hand gesture.

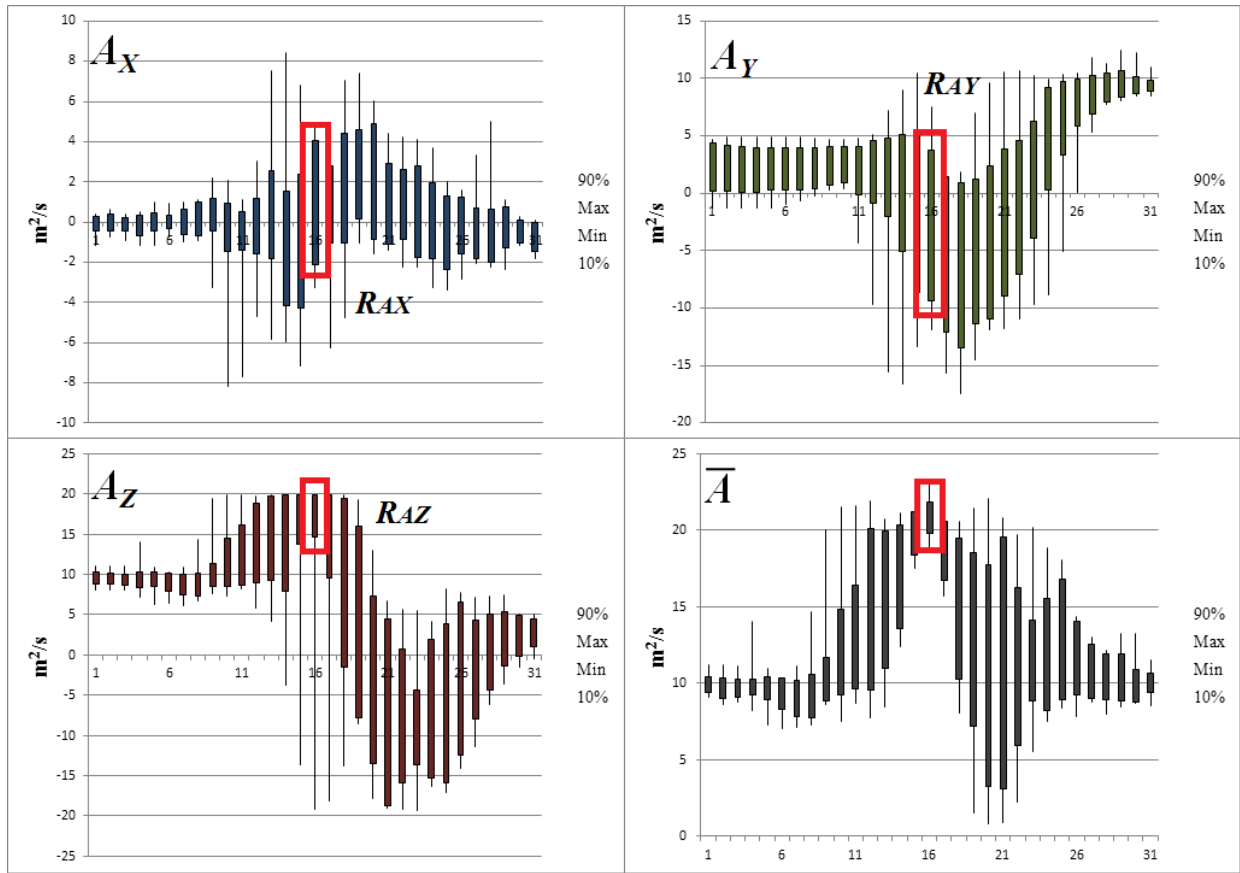


FIGURE 12. Acceleration data of “dragging” hand gesture.

significant advantage of this hand gesture detection approach is its simplicity, which means it requires little time to process the data from the smartphone accelerometer and detect the “dragging” hand gesture accurately. This feature is important to the design of Smart Signage system, as will be elaborated in the next section.

### III. PERFORMANCE ANALYSIS

One design challenge of cyber-physical systems is that they involve processes in the physical world that are not entirely predictable, which might cost systems’ stability or performance [20]. In the context of interactive display systems, the unpredictability of users’ behaviors might compromise users’ QoE, which is equivalent to the system performance. It is important to quantitatively categorize the user experience incorporating the interaction process. For Smart Signage, the interaction delay perceived by the user is identified as the most important user experience parameter. The performance of the system will first be analyzed with a single digital display, and then extended to the case with multiple displays.

#### A. PROPOSED SYSTEM WITH A SINGLE DISPLAY

The timing profile of the system with one display is shown in Fig. 13. The signage device transmit  $K$  “draggable” files in a

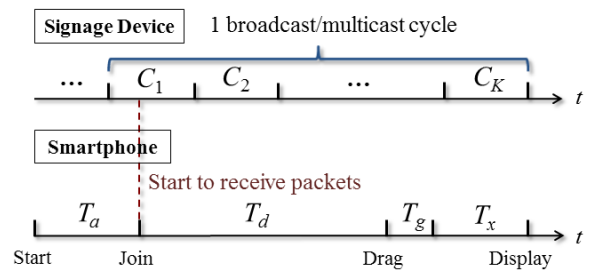


FIGURE 13. Timing profile of signage device and smartphone.

round-robin fashion and the smartphone goes through a process characterized by a series of time intervals. *Association time*  $T_a$  is the duration of joining the B/M group. *Decision time*  $T_d$  is the time interval between smartphone successfully joins the B/M group and the user performs a dragging gesture. *Gesture detection time*  $T_g$  is the time required for the smartphone to recognize the dragging gesture. *Transmission time*  $T_x$  is the time interval between a dragging gesture is confirmed and the intended content is displayed on the smartphone.

The performance of Smart Signage is characterized by *response time*  $T_r$ , which is the interaction delay perceived by the user.  $T_a$  is excluded from  $T_r$  as association process is prior to interaction and the duration is unnoticeable because

of skillful implementation, which will be explained in the next section. As  $T_d$  is primarily determined by the user, it is also not counted towards  $T_r$ . Thus  $T_r$  is defined as:

$$T_r = T_g + T_x. \quad (3)$$

In Eq. (3),  $T_g$  depends on the complexity of the gestures, which is small as only one simple gesture is in this system, and  $T_x$  depends on  $T_d$ . To be more specific, if  $T_d$  is larger than a B/M cycle, then  $T_x$  is 0 as the intended “draggable” file has already been buffered, otherwise,  $T_x$  is a function of  $T_d$ . To compute  $T_x$ , some assumptions are made to simplify the problem.

- 1) As the serving rate of the wireless router is much larger than the rate of the signage devices, an ideal channel is assumed that there is no packet loss.
- 2) Each “draggable” file has the same size and is uniformly divided into the same number of datagrams to be transmitted as B/M packets, and the time needed to send out one complete “draggable” file (*file transfer time*) is denoted as  $T_f$ .
- 3) The B/M packets of each “draggable” file will be repeatedly transmitted again in a round-robin fashion after the turns of sending the packets of other “draggable” files are done.
- 4)  $T_c = K \times T_f$  is the time of one complete B/M cycle and the analysis will only focus on  $T_c$  after the user has successfully joined the B/M group.
- 5) Only  $T_d$  that falls into the range of  $0 \leq T_d < T_c$  would be considered.
- 6) The user successfully joining a B/M group is a random process, which is uniformly distributed over one complete B/M cycle  $T_c$ .

Fig. 14 shows the diagram of a simple example to calculate  $T_x$ , where  $K = 3$ ,  $T_f < T_d < 2T_f$  and the user is interested in  $C_3$ . The formulation is divided into 3 cases according to  $t$ , defined as the time interval between the time point when a complete B/M cycle starts and the time point when the smartphone has successfully joined the B/M group, where  $0 \leq t < T_c$ .

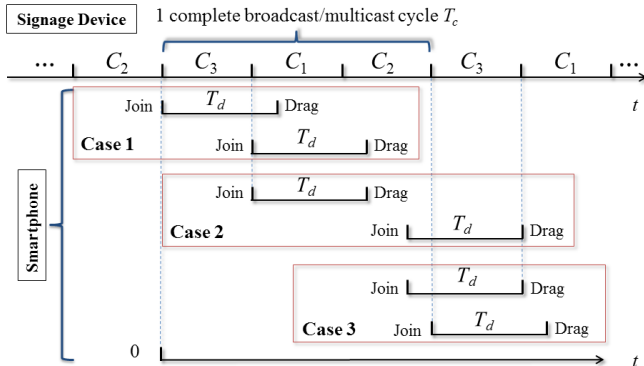


FIGURE 14. An example to calculate  $T_x$  with  $K = 3$ ,  $T_f < T_d < 2T_f$  and  $C_3$  is interested.

- 1) *Case 1*:  $0 \leq t < T_f$ , as only part of  $C_3$  is received,  $T_g + T_x = T_c - T_d$ .
- 2) *Case 2*:  $T_f \leq t < T_c + T_f - T_d$ , where the user has to wait until  $C_3$  has been received, so  $T_g + T_x = T_c - T_d - (t - T_f)$ .
- 3) *Case 3*:  $T_c + T_f - T_d \leq t < T_c$ , as all the packets of  $C_3$  are received before the “dragging” event,  $T_x = 0$ .

One thing to notice in this example is that although the user is interested in  $C_3$ , the results will be the same for any content,  $C_k$ , as one complete B/M cycle is considered. As the gesture detection time  $T_g$  is constant and negligibly small because of some skillful implementation discussed before,  $T_g = 0$  is assumed in the following analysis further simplify the formulation problem. In this case, the transmission time  $T_x$  can be expressed as:

$$T_x(t) = \begin{cases} T_c - T_d & \text{for } 0 \leq t < T_f; \\ T_c + T_f - T_d - t & \text{for } T_f \leq t < T_c + T_f - T_d; \\ 0 & \text{for } T_c + T_f - T_d \leq t < T_c. \end{cases} \quad (4)$$

Eq. (4) is valid for  $K \geq 2$  and  $T_f \leq T_d < T_c$ , which can be verified in a similar manner. In terms of  $K \geq 2$  and  $0 \leq T_d \leq T_f$ ,

$$T_x(t) = \begin{cases} T_c - T_d & \text{for } 0 \leq t < T_f; \\ T_c + T_f - T_d - t & \text{for } T_f \leq t < T_c. \end{cases} \quad (5)$$

Eq. (4) and (5) are for the situation that  $K \geq 2$ . However, there could be only one content on the signage display (i.e.,  $K = 1$ ), hence the transmission time is:

$$T_x(t) = T_f - T_d \text{ for } K = 1. \quad (6)$$

As assumed earlier that a user successfully joining the B/M group, or equivalently  $t$ , is uniformly distributed over one complete B/M cycle  $T_c$ , taking the expectation of  $T_x$  with respect to  $t$  will give an *expected transmission time*  $\overline{T_x}$  experienced by the user as:

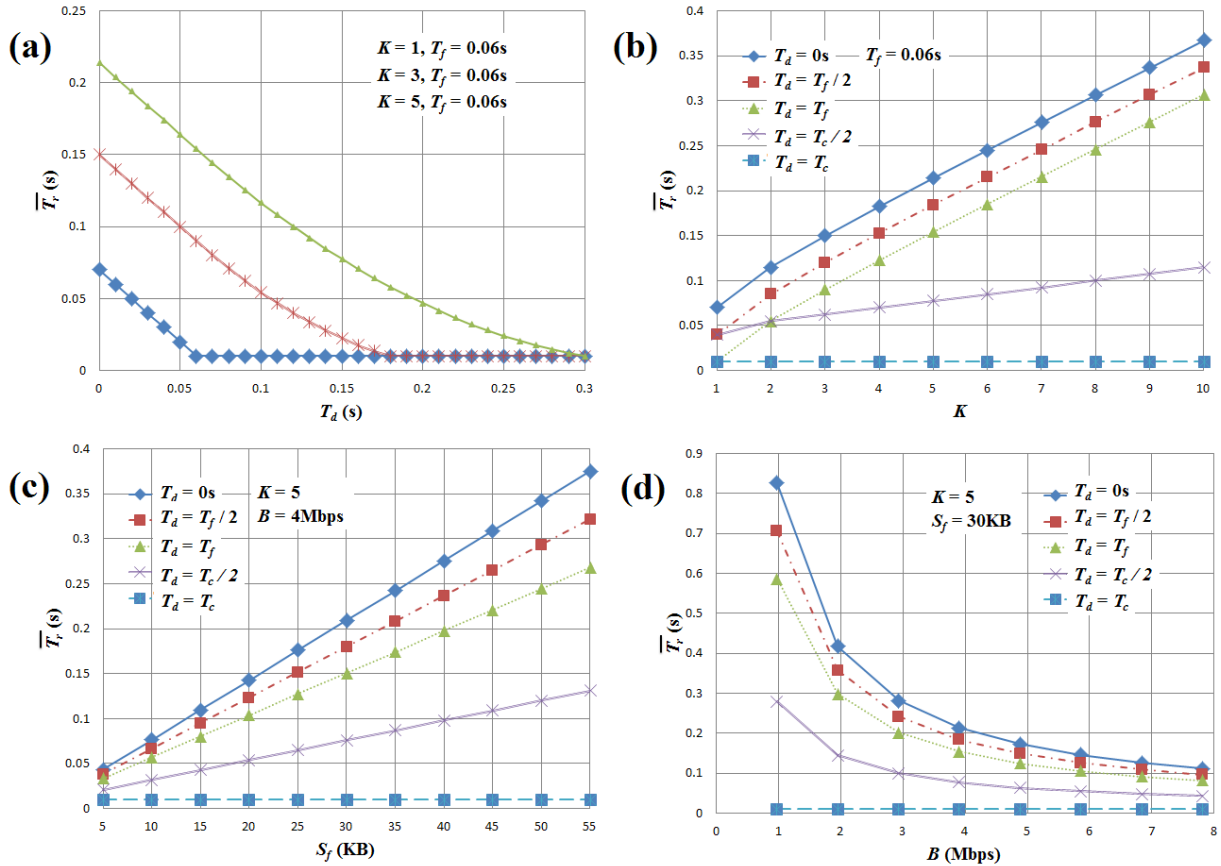
$$\overline{T_x} = \begin{cases} \left[ (T_c - T_d)T_f + \frac{1}{2}(T_c - T_f)(T_c + T_f - 2T_d) \right] / T_c & \text{for } 0 \leq T_d < T_f \text{ and } K \geq 2; \\ \left[ (T_c - T_d)T_f + \frac{1}{2}(T_c - T_d)^2 \right] / T_c & \text{for } T_f \leq T_d < T_c \text{ and } K \geq 2; \\ T_f - T_d & \text{for } K = 1. \end{cases} \quad (7)$$

Hence, the *expected response time*  $\overline{T_r}$  is:

$$\overline{T_r} = T_g + \overline{T_x}. \quad (8)$$

Eq. (7) shows that  $\overline{T_x}$  is a function of  $K$ ,  $T_f$ , and  $T_d$ . Note that  $T_f = S_f/B$ , where  $S_f$  is the size of the “draggable” file and  $B$  is the bit rate of the channel. So the expected response time of the system  $\overline{T_r}$  is a function of  $T_g$ ,  $N$ ,  $S_f$ ,  $B$  and  $T_d$ . As  $T_g$  is small, it is fixed to be 0.01s in the following numerical analysis.

Fig. 15(a) shows the plot of  $\overline{T_r}$  against  $T_d$ . As larger  $T_d$  means a longer time for the smartphone to buffer the “draggable” files,  $\overline{T_r}$  will decrease as  $T_d$  grows.  $\overline{T_r}$  is plotted



**FIGURE 15.** Expected response time  $\bar{T}_r$ : (a) against  $T_d$ ; (b) against  $K$ ; (c) against  $S_f$ ; (d) against  $B$ .

against  $T_d$  under the condition of  $K = 1, 3$  and  $5$ . Note that  $\bar{T}_r = T_g$  after  $T_d = T_c$ , as all the “draggable” files are buffered after  $T_d = T_c$ , such that  $\bar{T}_x$  will be  $0s$  afterwards.  $T_f$  is fixed at  $0.06s$  as the “draggable” file size  $S_f$  is set as  $30KB$  ( $240 \times 320$  JPEG image with the resolution quality sufficient for the application) and the bit rate  $B$  is assumed to be  $4Mbps$ . The plot shows that the decision time  $T_d$  required to achieve small  $\bar{T}_r$  is also small. When a user discovers an interesting content and performs a “dragging” hand gesture, the user will experience almost no delay before the content is showed on the smartphone. Even for the worst case, the expected response time,  $\bar{T}_r$ , will be less than 1 second.

Fig. 15(b) shows the plot of  $\bar{T}_r$  against  $K$ . A range of values of  $T_d$  were used in this numerical analysis, which are  $T_d = 0s$ ,  $T_d = T_f/2$ ,  $T_d = T_f$ ,  $T_d = T_c/2$  and  $T_d = T_c$ . These are boundary cases and turning points as well as two representative points in Eq. (7), such that the boundary cases of  $\bar{T}_r$  can be shown in the plot as well as some representative curves in between.  $T_f$  is fixed at  $0.06s$ . As observed from the plot, for each value of  $K$  from 1 to 10, the lower bound of  $\bar{T}_r$  is  $T_g$ , and the upper bound is approximately proportional to  $K$ . In overall, this system could give a reasonable expected response time,  $\bar{T}_r$ , for this range of  $K$ . Requiring a large number of contents in the system is not realistically needed

in the real-world deployments for advertising signages. If  $K = 5$ , the implication is that the user can freely “drag” up to 5 different contents, while the system can still maintain  $\bar{T}_r$  below 1 second.

Fig. 15(c) shows the plot of  $\bar{T}_r$  against  $S_f$ . The same values of  $T_d$  used in Fig. 15(b) are used in this plot.  $K$  is fixed at 5 and  $B$  is fixed at  $4Mbps$ . As observed from the plot, the lower bound of  $\bar{T}_r$  is  $T_g$ , and the upper bound is linear with  $S_f$ . It is not ideal to use a large and non-constant file size,  $S_f$ , if  $\bar{T}_r$  is needed to be a constantly small. On the other hand, it is not necessary to use a large  $S_f$  as the resolution of the smartphone display is limited. However, the size of the “draggable” file cannot be reduced too much, otherwise the user experience will be compromised. As a result, a  $240 \times 320$  JPEG image (i.e., approximately  $30KB$  in size) allows the user to “drag” a content with good quality, while still experiences a small response time below 1 second.

Fig. 15(d) shows the plot of  $\bar{T}_r$  against  $B$ . The same values of  $T_d$  used in Fig. 15(b) are used in this plot.  $K$  is fixed at 5 and  $S_f$  is fixed at  $30KB$ . It is observed from the plot that the lower bound of  $\bar{T}_r$  is  $T_a$  and upper bound of  $\bar{T}_r$  is inversely proportional to  $B$ . Although further increasing  $B$  will reduce  $\bar{T}_r$ , the gain will diminish as  $B$  becomes larger. Hence,  $B = 4Mbps$  is adequate to evaluate the real implementations



for achieving relatively small  $\bar{T}_r$ . Considering a non-ideal channel with effective bit rate  $B$ , Fig. 15(d) can be interpreted as a larger error probability leading to a lower effective bit rate and thus a larger  $\bar{T}_r$ . As observed from the plot,  $B = 4\text{Mbps}$  is close to an effective bit rate of Wi-Fi in common situations, it is sufficient to achieve small  $\bar{T}_r$  below 1 second in the real implementations using Wi-Fi.

### B. PROPOSED SYSTEM WITH MULTIPLE DISPLAYS

To determine the expected response time  $\bar{T}_r$  of the Smart Signage system with  $N$  displays, several further assumptions were made:

- 1) Each sign is assumed to have the same number of “draggable” files  $K$  and each content file is of the same size  $S_f$ . The time required to transmit one content file is denoted as  $T_f$ , which equals to  $S_f/B$ , where  $B$  is the bandwidth;
- 2) As shown in Fig. 16, the wireless router serves the packets from the signage devices in a content file based round-robin fashion ( $C(n, k)$  denotes the  $k$ th content file of sign  $S_n$ ). The time required to send all the content files is denoted as one extended B/M cycle,  $T'_c$ ;
- 3) As shown in Fig. 16, a user could starts the application on smartphone at any time instance  $t$  inside a B/M cycle, so  $t$  is assumed to be uniformly distributed over  $T'_c$ .

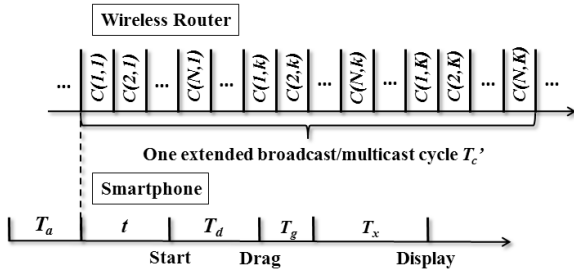


FIGURE 16. Packets arrival on wireless router and timing profile on smartphone.

To be more elaborate about Fig. 16, it shows an example in which the wireless router transmits the top-of-the-line content files of all the available displays first, and repeat the sequence for  $K - 1$  times for the rest of the content files within a B/M cycle. As the definition suggested,  $T'_c$  can be expressed as

$$T'_c = NKT_f. \quad (9)$$

Apply the results obtained in the last subsection about the transmission time  $T_x$ , for  $0 \leq T_d \leq T_f$ ,

$$T'_x(t) = \begin{cases} T'_c - T_d & \text{for } 0 \leq t < T_f \\ T'_c + T_f - T_d - t & \text{for } T_f \leq t < T'_c \end{cases}, \quad (10)$$

for  $T_f \leq T_d < T'_c$ ,

$$T'_x(t) = \begin{cases} T'_c - T_d & \text{for } 0 \leq t < T_f \\ T'_c + T_f - T_d - t & \text{for } T_f \leq t < T'_c + T_f - T_d \\ 0 & \text{for } T'_c + T_f - T_d \leq t < T'_c \end{cases}, \quad (11)$$

and for  $T_d \geq T'_c$ , as all the content files have already been buffered in local storage,

$$T'_x(t) = 0. \quad (12)$$

As  $t$  is uniformly distributed over one B/M cycle  $T'_c$ , so the expected transmission time is

$$\bar{T}'_x = \begin{cases} \left[ (T'_c - T_d)T_f + \frac{1}{2}(T'_c - T_f)(T'_c + T_f - 2T_d) \right] / T'_c & \text{for } 0 \leq T_d < T_f \\ \left[ (T'_c - T_d)T_f + \frac{1}{2}(T'_c - T_d)^2 \right] / T'_c & \text{for } T_f \leq T_d < T'_c \\ 0 & \text{for } T_d \geq T'_c \end{cases}, \quad (13)$$

which means,

$$\bar{T}'_r = T_g + \bar{T}'_x. \quad (14)$$

Suitable values for some of the parameters were explored in the previous subsection, where  $T_g$  is  $0.01s$ , the number of “draggable” files on each sign  $K$  is 5 and the bandwidth of the wireless router  $B$  is  $4\text{Mbps}$ . The relationship of expected response time  $\bar{T}_r$  and number of displays  $N$  were studied for two file sizes,  $S_f = 30\text{KB}$  and  $S_f = 100\text{KB}$ . Fig. 17 shows that, with  $S_f = 30\text{KB}$ ,  $\bar{T}_r$  is below 1.5 seconds for  $N$  from 1 to 10 even for the worst case where  $T_d = 0$ , which means medium quality images can be transferred effectively with small expected response time using the Smart Signage system with multiple digital displays.

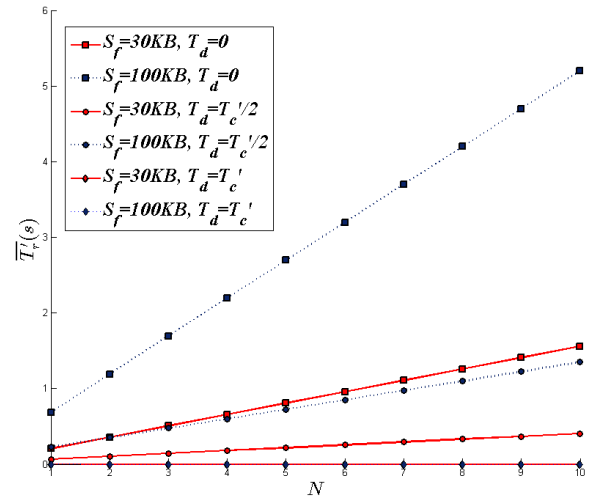


FIGURE 17. Plot of  $\bar{T}'_r$  against  $N$  with  $T_d = 0$ ,  $T'_c/2$  and  $T'_c$  ( $S_f = 30\text{KB}$  for solid lines and  $S_f = 100\text{KB}$  for dashed lines).

## IV. IMPLEMENTATION AND EXPERIMENTATION

### A. SYSTEM IMPLEMENTATION

Fig. 18 shows the implementation of Smart Signage system, which consists of a LCD TV as the signage display, a programmed VxWorks-based embedded media playback system as the signage device, and an IP-multicast enabled Wi-Fi wireless router. The wireless router is used by the signage device to multicast packets to smartphones and access

Internet. Software applications are implemented on Android and iOS smartphones. For Android platform, the association process of connecting to SSID and joining multicast group is handled by software automatically. The resulting association time  $T_a$  is within seconds. Considering the fact that this process occurs only once when the program starts, users will not experience  $T_a$  as delay at all. Fig. 18(b) shows using an Android smartphone to “drag” the content.

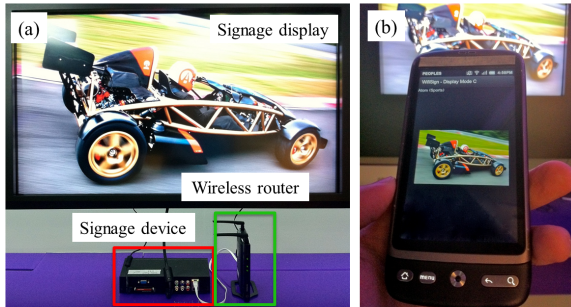


FIGURE 18. Implementation of the system.

To test the system working with multiple digital displays, the proposed system was implemented in a lab environment with ten LCD panels served as signage displays, which is shown in Fig. 19.



FIGURE 19. Lab environment with multiple signage displays: (a) north side with two displays; (b) east side with two displays; (c) south side with three displays; (d) center with three displays.

## B. IMPLEMENTATION OF THE “DRAGGING” HAND GESTURE

Fig. 20 shows the photographs of the 3 steps of a “dragging” gesture in the real implementation. A vertical levitation of the smartphone is observed during the process. Evoked by the “dragging” event, the corresponding visual responses to these 3 steps are implemented accordingly: Step 1: Before a user performs a “dragging” gesture, the display on the smartphone shows “Drag what you like”; Step 2: After a “dragging” event is detected, a successful “dragged” content is displayed on the smartphone; Step 3: Finally, the user performs follow-up actions to store and use or even discard the “dragged” content.

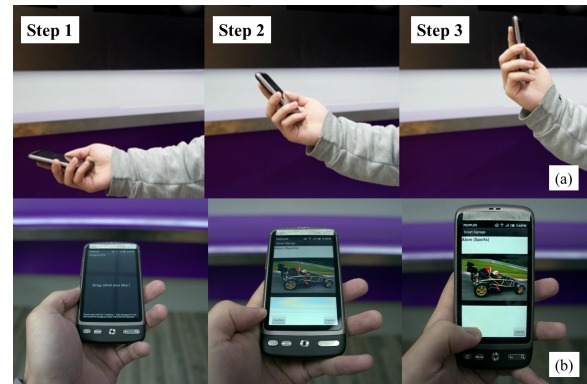


FIGURE 20. Implementation: a) the 3 steps of the “dragging” gesture; b) the corresponding visual responses.

## C. EXPERIMENT WITH MULTIPLE USERS

In order to test the system implementation, a series of experiments are done. In one of representative experiments, 30 primary school students were invited to the laboratory to experiment with the proposed cyber-physical system. Smartphones with different brands and OSs are installed with the proposed client-side mobile applications, which are distributed to these students and let them to “drag” the content on a display while they are walking around in the space of the laboratory. Without any prior knowledge about the interaction modality, these students easily got themselves familiar with the system after seeing a quick demonstration of “dragging” a content from the display into the smartphone showed by a researcher.

In order to test the system implementation, a series of experiments are done. In one representative experiment, 30 primary school students were invited to the laboratory to experiment with the proposed cyber-physical draggable B/M media system. Smartphones manufactured by different companies and installed with customized application were distributed to these students. With almost no prior knowledge of the principle of the interaction modality, these 30 students easily got themselves familiar with the system after following only one-minute demonstration of “dragging” the content from the signage display to their smartphones. The actual response time of the system is small enough that no students in this experiment reported experiencing delay during interaction. Fig. 21 shows the students with the smartphones who have successfully dragged the intended content to their devices simultaneously. This actually proved that Smart Signage successfully provides an intuitive interactive modality such that even primary school students can quickly learn. This also demonstrated that capability of the system to enable digital signages to simultaneously interact with multiple users while delivering satisfying user experience, even when the users are in mobile.

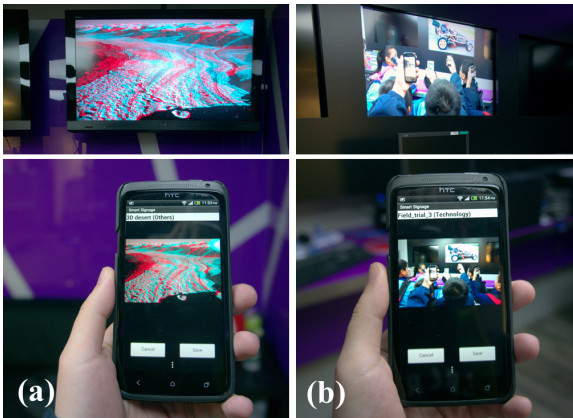
## D. EXPERIMENT WITH MULTIPLE DISPLAYS

Two representative setups of the signage displays were experimented in the lab environment to verify the feasibility of Smart Signage to accommodate multiple displays. Sixteen



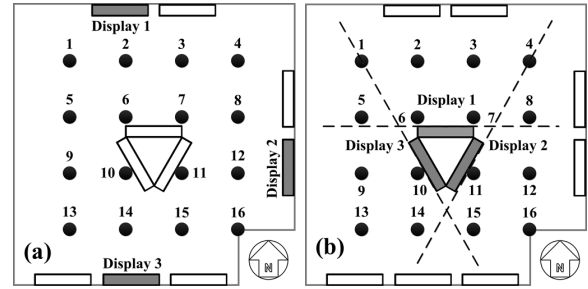
**FIGURE 21.** Smart Signage used by multiple primary school students.

equally spaced test points were selected inside the location to examine whether the proposed system can correctly differentiate these displays by their orientations. For any one of the sixteen test points, if a user has a clear view of the target signage display, the user should be able to “drag” the content on the display by pointing his/her smartphone at the signage display, otherwise, an error occurs. An example is given in Fig. 22 where the user was standing at a one test point and had clear views of two signage displays with different orientations. In this example, the user pointed the smartphone at each of the signage display and correctly “dragged” the content on the display, which means no error occurs. If the user in the example pointed at the first signage display and “dragged” the content from the second one, an error is detected at this test point for the first signage.



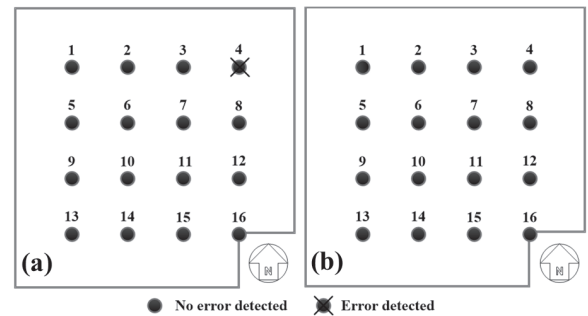
**FIGURE 22.** “Drag” content from one sign (a) and a second sign (b) with different signage display orientations.

The schematic of the first experiment is shown in Fig. 23(a), which involves three signage displays installed on north side, east side and south side of the lab walls. For all the sixteen test points, a user inside this location can have a clear view of all the three signage displays. So the user should experience no trouble “dragging” the correct content on these displays at any one of the test points. As shown in Fig. 24(a), no error is detected except for test point 4, an error was detected for Display 2. When a user tried to “drag” the



**FIGURE 23.** Schematics of the two experimental setups in the lab environment (rectangles are signage displays, where the solid ones are tested; black dots are equally spaced test points).

content from Display 2, the smartphone misinterpreted the target to be Display 3.



**FIGURE 24.** Results of the two experimental setups.

The schematic of the second experiment is shown in Fig. 23(b), which involves three signage displays installed at the center of the lab location. The dashed lines in the figure indicate the boundaries of the viewable area for each signage display. After excluding the test points that are on the boundaries or not in any viewable area of any display, test point 2, 3, 6 and 7 have clear view of Display 1, test point 11, 12, 15 and 16 have clear view of Display 2, and test point 9, 10, 13, 14 have clear view of Display 3. So a user should correctly “drag” the content on the signage display if he/she is standing at the test points inside the corresponding display’s viewable area. As shown in Fig. 24(b), no error was detected throughout this experiment.

## V. CONCLUSION

In this paper, Smart Signage, a “draggable” cyber-physical broadcast/multicast (B/M) media system is proposed in this paper. With a novel cyber-physical B/M protocol that synchronizes the content on the digital displays with the smartphones, it supports one-to-many interaction by allowing multiple users acquiring content from one display with a “dragging” hand gesture. With the embedded display orientation information in the same protocol, Smart Signage supports many-to-many interaction by allowing users obtaining content from the digital display they are pointing at with their smartphones. Users’ QoE which is characterized by the response time, is carefully studied in this paper to guarantee the performance of this cyber-physical interactive display sys-



tem. The system provides a new interactive modality for the users to obtain interested information from digital displays for advertising purpose, which has the advantages of scalability to engage a large number of users, more flexible interaction range, more intuitive way of interaction, less visual compromise on the presentation of the advertisement, higher throughput of data transmission and supporting interaction with mobile users.

## REFERENCES

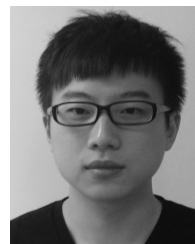
- [1] (2009). *Weight Watchers Advertisement* [Online]. Available: <http://gcms.weightwatchers.com>
- [2] T. Vajk, P. Coulton, W. Bamford, and R. Edwards, "Using a mobile phone as a 'wii-like' controller for playing games on a large public display," *Int. J. Comput. Games Technol.*, vol. 2008, pp. 1–6, Jan. 2008.
- [3] N. Davies, A. Friday, P. Newman, S. Rutledge, and O. Storz, "Using bluetooth device names to support interaction in smart environments," in *Proc. 7th Int. Conf. Mobile Syst., Appl., Services*, 2009, pp. 151–164.
- [4] K. Cheverst, A. Dix, D. Fitton, C. Kray, M. Rouncefield, C. Sas, G. Saslis-Lagoudakis, and J. G. Sheridan, "Exploring bluetooth based mobile phone interaction with the hermes photo display," in *Proc. 7th Int. Conf. Human Comput. Interact. Mobile Devices Services*, 2005, pp. 47–54.
- [5] A. S. Shirazi, C. Winkler, and A. Schmidt, "Flashlight interaction: A study on mobile phone interaction techniques with large displays," in *Proc. 11th Int. Conf. Human-Comput. Interact. Mobile Devices Services*, 2009, pp. 93:1–93:2.
- [6] D. Schmidt, F. Chehimi, E. Rukzio, and H. Gellersen, "Phonetouch: A technique for direct phone interaction on surfaces," in *Proc. Annu. ACM Symp. UIST*, 2010, pp. 13–16.
- [7] R. Ballagas, M. Rohs, and J. G. Sheridan, "Sweep and point and shoot: Phonetouch-based interactions for large public displays," in *Proc. Extended Abstracts Human Factors Comput. Syst.*, 2005, pp. 1200–1203.
- [8] N. Pears, D. Jackson, and P. Olivier, "Smart phone interaction with registered displays," *IEEE Pervas. Comput.*, vol. 8, no. 2, pp. 14–21, Apr./Jun. 2009.
- [9] (2008). *Fox Quick Response Code Advertisement* [Online]. Available: <http://www.amplify.com.au/blog/tag/coding/>
- [10] (2012). *U-Tie* [Online]. Available: <http://www.u-tie.com.hk/front/en/home>
- [11] K. Seewoonauth, E. Rukzio, R. Hardy, and P. Holleis, "Touch & connect and touch & select: Interacting with a computer by touching it with a mobile phone," in *Proc. 11th Int. Conf. Human-Comput. Interact. Mobile Devices Services*, 2009, pp. 36:1–36:9.
- [12] R. Hardy, E. Rukzio, M. Wagner, and M. Paolucci, "Exploring expressive NFC-based mobile phone interaction with large dynamic displays," in *Proc. IEEE Comput. Soc. 1st Int. Workshop NFC*, Sep. 2009, pp. 36–41.
- [13] R. Dachsel and R. Buchholz, "Natural throw and tilt interaction between mobile phones and distant displays," in *Proc. 27th Int. Conf. Extended Abstracts Human Factors Comput. Syst.*, 2009, pp. 3253–3258.
- [14] J. Scheible, T. Ojala, and P. Coulton, "Mobotoss: A novel gesture based interface for creating and sharing mobile multimedia art on large public displays," in *Proc. 16th ACM Int. Conf. Multimedia*, 2008, pp. 957–960.
- [15] RPA. (2011). *Mobile Mouse for the Iphone, Ipad, & Ipad*, Baltimore, MD, USA [Online]. Available: <http://www.mobilemouse.com/overview.html>
- [16] A. Bellucci, A. Malizia, P. Diaz, and I. Aedo, "Human-display interaction technology: Emerging remote interfaces for pervasive display environments," *IEEE Pervas. Comput.*, vol. 9, no. 2, pp. 72–76, Apr. 2010.
- [17] B. Kemp, A. J. Janssen, and B. van der Kamp, "Body position can be monitored in 3D using miniature accelerometers and earth-magnetic field sensors," *Electroencephalogr. Clinical Neurophysiol./Electromyogr. Motor Control*, vol. 109, no. 6, pp. 484–488, Dec. 1998.
- [18] A. Waluyo, B. Srinivasan, and D. Taniar, "Optimal broadcast channel for data dissemination in mobile database environment," in *Advanced Parallel Processing Technologies* (Lecture Notes in Computer Science), vol. 2834. Berlin, Germany: Springer-Verlag, 2003, pp. 655–664.
- [19] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Trans. Syst., Man, Cybern., C, Appl. Rev.*, vol. 37, no. 3, pp. 311–324, May 2007.
- [20] E. Lee, "Cyber physical systems: Design challenges," in *Proc. 11th IEEE ISORC*, May 2008, pp. 363–369.



**JAMES SHE** is an Assistant Professor in the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), Hong Kong, and a Visiting Research Fellow with the University of Cambridge, Cambridge, U.K. He is the founding director of Asia's first social media laboratory, HKUST-NIE Social Media Laboratory, and spearheads multidisciplinary research and innovation in cyber-physical social media systems, viral media analytics, and mobile media broadcast systems. Celebrated as a thought leader in new media and emerging cyber-physical societies, he is a member of the World Economic Forums Global Agenda Council (Social Media) and joins other government and business leaders to develop solutions to key social media issues on the global agenda.



**JON CROWCROFT** has been the Marconi Professor of communications systems in the Computer Laboratory since October 2001. He has worked in the area of Internet support for multimedia communications for over 30 years. His main topics of interest have been scalable multicast routing, practical approaches to traffic management, and the design of deployable end-to-end protocols. His current active research areas are opportunistic communications, social networks, and techniques and algorithms to scale infrastructure-free mobile systems. He leans towards a "build and learn" paradigm for research. He received the Degree in physics from Trinity College, University of Cambridge, Cambridge, U.K., in 1979, and the M.Sc. degree in computing in 1981 and the Ph.D. degree in 1993 from University College London, London, U.K. He is a Fellow of the Royal Society, ACM, the British Computer Society, the IET, and the Royal Academy of Engineering. He has published a few books based on learning materials.



**HAO FU** received the B.Eng. degree from the Electronic and Computer Engineering Department, Hong Kong University of Science and Technology, Hong Kong, in 2011. He is currently with HKUSTNIE Social Media Laboratory, and is currently pursuing the M.Phil. degree from the Electronic and Computer Engineering Department, Hong Kong University of Science and Technology. His research interests include cyber-physical systems for multimedia applications, mobile and pervasive computing, and online social networks analysis and applications.



**PIN-HAN HO** received the B.Sc. and M.Sc. degrees from the Electrical Engineering Department, National Taiwan University, Taipei, Taiwan, in 1993 and 1995, respectively, and the Ph.D. degree from Queens University at Kingston, Kingston, U.K., in 2002. He is currently an Associate Professor in the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He is the author or co-author of more than 250 refereed technical papers, several book chapters, and the co-author of a book *Optical Networking and Survivability*. His current research interests cover a wide range of topics in broadband wired and wireless communication networks, including survivable network design, wireless metropolitan area networks such as the IEEE 802.16 networks, fiber-wireless network integration, and network security. He is the recipient of the Distinguished Research Excellent Award in the Electronic and Computer Engineering Department of University of Waterloo, the Early Researcher Award (Premier Research Excellence Award) in 2005, the Best Paper Award in SPECTS'02, ICC'05 Optical Networking Symposium, and ICC'07 Security and Wireless Communications Symposium, and the Outstanding Paper Award in HPSR'02.