The crowd in the cloud: moving beyond traditional boundaries for large scale experiences in the cloud

Adam Roughton[1] **John Downs**[1,2] **Beryl Plimmer**[1] **Ian Warren**[1]

[1] Department of Computer Science
University of Auckland, New Zealand
Email: arou012@aucklanduni.ac.nz {beryl|ian-w}@cs.auckland.ac.nz

[2] Department of Information Systems
University of Melbourne, Australia
Email: j.downs@pgrad.unimelb.edu.au

Abstract

In this paper we propose a taxonomy for crowd based interaction paradigms, and categorise the literature according to this taxonomy. The conventional definition of crowds needs to be reconsidered in the light of advances in communication technology such as smart phones and cloud based infrastructures. We have extended the definition to encompass virtual dispersed crowds by considering the core components of crowd based activities. We found that much of the existing work offers simplistic reactive control in exchange for economical, highly synchronous, co-operative activities. We argue that the same co-operative component and economy can be obtained with rich reflective control. By combining the cloud, smart phones, and tools, this gap can be exploited to create a new class of rich, thought provoking, economical, crowd computing.

Keywords: Crowd computer interaction, crowd computing, crowd and the cloud, collective behaviour

1 Introduction

Humans are an inherently community-based species; we associate with nations, societies, and clubs. Our interaction with society at large is often at times of shared activity: events like the Olympics are able to engage the passions of entire nations. There exists little technology that attempts to engage communities in crowd activity. Work in the emerging field of crowd computer interaction attempts to address this shortcoming.

Crowd computer interaction first came to prominence in 1991 during the SIGGRAPH Electronic Theatre show with the deployment of Cinematrix Interactive Entertainment System (Carpenter, 1993). Each audience member was given a double sided paddle with a green face on one side and a red one on the other; cameras tracked the paddles in real-time allowing for collaborative audience activities. In this way the audience was able to collectively control a single paddle in a game of pong, vote on an issue, or move through a maze (Maynes-Aminzade et al., 2002; Bregler et al., 2005). The system was invented to get around the expense and the required preparedness of wiring seats with controllers, or the need to hand out expensive wireless controllers to audience members (Carpenter, 1993). Much of the subsequent work in this area has built on this foundation of throwaway sensors deployed to a co-located crowd (or in some work no controller at all) and a large shared screen for feedback.

Crowds present a unique problem space for the deployment of technology. This is because of the organisation and composition of individuals within the crowd, and the sheer scale presented by the possible numbers of participants. The organisational structure of the crowd takes on a looser form than groups (Turner and Killian, 1972): crowds are far more adhoc in their nature, and in most cases there is no clear leadership. Crowd membership is not as stringently regulated as groups with participation often open to anyone who is able to take part. The relaxed form of membership and the large number of people leads to a collection of many different cultures and identities. Within the larger crowd aggregate are many small sub-crowds and groups with their own idiosyncrasies and traditons (Reicher, 2002). Behaviours appropriate to the traditions of one sub-crowd may not be appropriate for another. How the crowd works as a whole to work around these differences, or how it might attenuate or exaggerate issues remains a new area of inquiry for technology solutions deployed into this arena.

In this paper we considered how two new technologies, smart phones and cloud computing, can change crowd participation.

The recent proliferation of smart phones equipped with accelerometers, compasses, microphones, and gyroscopes eliminates the need to build, develop, or distribute sensors. This paired with the rich interfaces provided by the devices opens up the possibility of richer crowd experiences. Application marketplaces have provided the ability to quickly distribute applications to client phones, eliminating the original problem of sensor deployment cost.

Cloud computing provides the second piece of the puzzle: the cloud computing paradigm reduces entry barriers with reduced infrastructure costs and quick access to computing power and storage (Greengard, 2010). The potential benefits of cloud computing to crowd computing are vast. By their very nature, crowd applications benefit from the elasticity offered by the cloud (Owens, 2010): the ability to scale up a modest infrastructure for short durations at a very modest cost fits the typical crowd application scenario beautifully.

The combination of smart phones and cloud computing opens the possibility for rich, economical crowd computer interaction without the need for colocation.

This paper first proposes a new definition of crowds for the purpose of crowd computer applica-

Copyright ©2011, Australian Computer Society, Inc. This paper appeared at the 12th Australasian User Interface Conference (AUIC2011), Perth, Australia, January 2011. Conferences in Research and Practice in Information Technology (CRPIT), Vol. 117, Christof Lutteroth and Haifeng Shen, Ed. Reproduction for academic, not-for profit purposes permitted provided this text is included.

tions based on the developments in communication technologies. A taxonomy based on the key components of crowd activity is then presented, followed by a categorisation of applications in the literature. A discussion of the trends and gaps found in the categorisation follows before the conclusion of the paper.

2 Definitions

Some definitions are in order before a taxonomy can be approached.

First and foremost, what is meant by the 'crowd' needs to be defined. Formally, the American Psychology Association defines a crowd as "a sizable gathering of people who temporarily share a common focus and a single location" (VandenBos, 2007, pg 247). This definition is rather ambiguous for the purposes of designing applications for crowds. What is a 'sizable gathering'? How long is 'temporarily'? With the advent of technologies that support large numbers of people in a shared activity without spatial constraints (Twitter, 2010; Blizzard Entertainment, 2010), the boundaries of what has traditionally been deemed a crowd are challenged.

A novelty for research into crowd based interaction is the unique form of behaviour found in crowd based settings. Turner and Killian (1972) define *collective behaviour* as all those behaviours found in collectives that are not based on the traditional behaviour of society (Turner and Killian, 1972). In conventional settings an individual is subject to the social norms of society, the workplace, or teams: established operating procedures guide what behaviour is appropriate. They go on to suggest crowd behaviour is similar to that of groups in that members interact with one another with a sense that they constitute a unit; the behaviours of members are clearly influenced by the other members, and social norms are arrived at by all. Where crowd behaviour differs from group behaviour is the formation of these social norms. Unlike groups, crowds have a looser structure with regards to membership and leadership: member participation is unregulated by any existing structure, and leaders are those that the crowd chooses to follow at a given moment. While groups have procedures that are based on society at large, crowds have a far more spontaneous form of behaviour that stems from the looseness of structure; some norms may reflect those in society, others might outright reject them. There may in fact be competing social norms in a crowd. The interest of crowds for technology deployment are supporting crowd activities for large numbers of people with this collective behaviour.

For the purposes of this paper, a crowd is defined as a *large collective* of people under a *synchronous shared influence*, where each member is connected to one another via some *communication medium*, and all members have the *potential to participate* in the *collective activity*.

This definition raises points that need to be addressed:

- **collective activity**: refers to the defining activity of the crowd; viewed as the focal point for a given crowd. This is linked to both explicit membership identity (I am part of the crowd watching the All Blacks vs Wallabies game at Eden Park), and implicit membership identity (I am part of the crowd watching the All Blacks vs Wallabies game at Eden Park *singing the New Zealand national anthem*). Sub-crowds are often identified by specifying the activity with greater precision. In the preceding example, Australian supporters in the stadium crowd are often not able to participate in singing the anthem due to social pressure; they are hence not part of the singing the New Zealand national anthem sub-crowd, though they still spectate its output. They still retain membership in the greater stadium crowd along with those participating in the anthem.
- **large collective**: the crowd size has been defined ambiguously on purpose: existing work in this area has crowds ranging from less than a hundred participants through to the tens of thousands. What distinguishes crowds from traditional groups is the presence of collective behaviour. The size of the collective must therefore be large enough for this behaviour to be a property of the crowd. The modifier large also hints at the requirements of any potential system being able to cope with tens of thousands of members, while also being able to support smaller gatherings.
- **communication medium**: each crowd member is connected in some way to every other crowd member via some medium. In this way, each member is able to have some input into the collective activity. In co-located settings, this medium is most often made up of visual and aural channels: crowd members can be seen and heard by others in the crowd, and their participation forms a part of what is seen as the crowd.
- **synchronous shared influence**: the influence refers to the output of the collective activity; each member has an awareness of their role in the crowd and is affected in some way by the output of the crowd. All crowd members semantically share the same influence simultaneously, but perhaps via different mediums.
- **potential to participate**: all crowd members must be able to contribute input towards the activity's output; any choice to not actively participate is based solely on the individual's choice not to participate. Other limiting factors including physical (e.g. no means of participating), social (e.g. peer pressure), and temporal (e.g. outside of the activity's timeframe) limitations may prevent an individual from being considered part of the crowd.

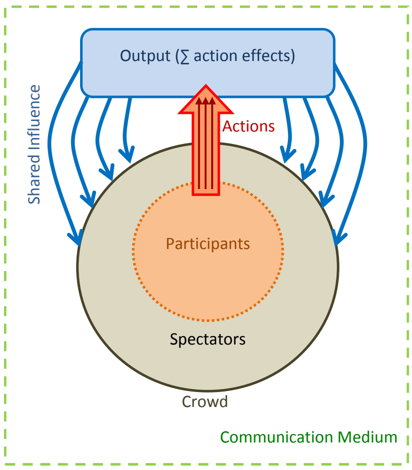

Figure 1: Components of a crowd activity.

The above definition specifies that those with the *potential* for participation are members of the crowd. Spectators of a crowd activity already satisfy the criteria for shared influence (spectators are influenced by the output). Hence, it is implied that all spectators of a crowd activity are also members of the crowd as long as they have the potential for involvement in the activity. This potential for participation implies that spectators share the communication medium of participants and so are able to input into the activity. Every crowd activity (collective activity) involves a set of *actions* made by participants. The complete *action set* defines the crowd activity. Each action has at least one consequential *action effect.* Figure 1 puts the components of a crowd activity together according to the above definition.

As an example, the well known stadium activity of the "Mexican wave" has two actions: standing up with one's arms outstretched, and remaining seated. Both of these actions are crucial to the recognisable activity: participants who are not part of the wave must remain seated until the wave reaches their part of the crowd, and stand only for the instant at which the wave reaches their seat section. The effect of each action is visible in the output of the activity, namely the movement of the wave around the stadium. All crowd members are able to spectate the output of the crowd by watching the participant constructed wave move around the stadium.

3 Taxonomy

A taxonomy provides a useful method of exploring the domain space: prior and ongoing work can be related to the domain as a whole. This is especially useful in the crowd computer interaction domain as there has been little work in the area of classification of applications in the domain space.

The taxonomy presented here builds on the observations made from existing work, with the help of sociological and psychological research into the phenomenon of crowds.

3.1 Intention

In classifying the interaction individuals have both with and within the crowd, an important distinction must be made between the actions and the associated intentions, due to the occurrence of collective intention in crowd activity.

In the case of a single individual, an intention leads neatly into an associated action. However, in the case of crowd or group intentionality, experience tells us that a shared intention comes into play. One proposition terms this shared intention 'collective intention' (Searle, 1990). Searle argues that a collective intention is not the mere result of individual intentions, but that it is a distinct primitive construct of its own. An individual specifies the intention with the form 'we intend' to do such and such, as opposed to 'I intend'. Searle further stipulates that 'I-intentions' then derive from the collective 'we-intentions': a collection of soccer players has the intention 'we intend to perform play X', with each player having an individual intention of the form 'we intend to perform play X by means of my action Y'.

Although this is still an ongoing discussion in philosophy, the concept of 'we-intentions' is useful in describing the intuition that collective intentionality manifests different behaviour in group settings than a mere collection of single individual intentions.

At the top level, crowd activities can be separated based on the intention makeup of the crowd during the activity:

- **Individual**: crowd members have many different I-intentions
- **Shared**: crowd members share the same intention

Activities that are made up from members of the crowd with different I-intentions (or in some cases we-intentions) are treated first. In this class of activity, individuals have their own goals and objectives. These objectives are performed inside the activity domain: members recognise themselves as actors within some common activity. Street crowds are an example of this form of activity that many of us are familiar with; as actors within the crowd, each person has their own intentions to get to their intended destinations, or perhaps even to just wander the streets. On entering the crowd, they enter into the shared influence of the crowd activity, namely the flow of the crowd. Whether or not they wish to be in the crowd, each actor recognises themselves as part of the crowd as they pass through it.

Activities of this type are by no means always made up incidentally in the same way as street crowds; participants may wish to engage in some shared activity by which the presence of others in the activity is required for the experience, but the intentions of participants is individual. Within a stadium crowd a member might wave a banner of support for their team: this action requires the presence of other crowd members for the action to have the intended effect, but does not require all crowd members have the same intention.

Shared intention activities are defined by crowd members sharing the same intention. The activity class can be further divided based on the type of the shared intention:

- *Co-operative* intentioned activities are those where every participant (or most participants) of the crowd have the same we-intention. These activities involve all participants working together to solve some common objective.

- *Separate* intentioned activities involve every participant having the same I-intention. An example of this class of activity would be a competitive pull on some shared resource. Voting systems also fall into this class; although individual members may vote for separate outcomes, the activity of voting is a shared I-intention. Activities of this class might also appear to be shared activities from observation: crowd technologies that aggregate the input of all participants to control some shared control (an example here would be the pong game from (Carpenter, 1993)), might result in individual crowd members having I-intentions (I intend to move the paddle left) if the crowd does not communicate within itself. However, if crowd members co-ordinate themselves through communication such as yelling, and cheering, the activity becomes a co-operative one.

3.2 Interaction setting

The interaction form describes the configuration of the actions contained within a single interaction with the system. A single interaction is defined as starting from the point of first input until the point of output presentation. Each interaction setting describes both the origin and the destination for the action(s) (Whitworth and Plimmer, 1997):

- **One-to-many (1:N)**: each interaction involves the input of a single crowd member with output (action effects) presented to many crowd members. In this way the action is amplified. An example is the activity of hitting a beach ball around a crowd: each strike of the beach ball is made by a single (or a marginally small subsection) of the crowd with the action effects of the raised ball visible by all.

- **Many-to-many (N:M)**: each interaction involves the input of many crowd members with output presented to many other crowd members. An example of this is crowd pong (Maynes-Aminzade et al., 2002), where all crowd members act together to control a single paddle. In this instance a single interaction involves the input of all participants, with the output directed to all crowd members.

- **Many-to-one (N:1)**: for each interaction, the input of many is presented to a single crowd member (or a very small subset of the crowd). The action effects are therefore aggregated. This can be seen in stadiums where a crowd member is singled out on a large display, and adjusts their behaviour based on the feedback of the crowd.

3.3 Interaction style

Describes the style of interaction encouraged or required by the activity.

- **Reflective**: the input into the activity is based on reflection and real creative input; the activity encourages participants to reflect before they contribute to the output. Reflective actions allow purposeful and creative input into the system: members might use strategy to plan their action, or the input might require creativity in expression or content. An example of this can be found in the use of Twitter during events: each 'tweet' requires careful reflection to properly express the thoughts of the participant before the expression is presented to the crowd.

- **Reactive**: the input is based on quick reaction to some stimuli: crowd members are encouraged to act quickly for their contribution to the output. Reactive actions are based on coordination and speed: members need to quickly process the stimuli and react accordingly. Activities like crowd pong (Maynes-Aminzade et al., 2002) require that all participants lean in time to control the onscreen paddle; participants must time their lean carefully to move the paddle into the path of the oncoming ball.

3.4 Social Exposure

The degree to which an individual in the crowd is 'socially exposed' depends on three factors:

- **Visibility of the action effects**: this factor describes how visible the action effect is to other members of the crowd. Highly visible effects are those that are accessible by all members, while effects with low visibility are accessible by only a few members.

- **Strength of the association between an individual and their action effects**: a strong association allows crowd members who can see the effect to be able to always find the individual who performed the action. A weak association requires some effort to be made by crowd members who can perceive the effect, while no association means that an individual can never be linked to their actions.

- **Duration of the action effects**: a long duration means that the effects of the action are visible for a long period of time, while a short duration means that the effects of an action are quickly removed from the activity.

3.5 Crowd member connectivity

This describes the 'closeness' of crowd members in terms of communication channels: crowd members are close if they share a large amount of information amongst themselves. The information relayed by one member to another depends on three factors related to the communication channels between them (Whitworth and Plimmer, 1997):

- **Channel type**: describes the media type of the channel. Primary channels relate to the senses; these include visual, audial, and tactual amongst others. Symbology is a secondary channel that relates to the symbology found in communication such as language.

- **Number of channels**: the number of communication channels in the media space.

- **Channel bandwidth**: describes the information content capacity and speed of the channel. High bandwidth channels convey a large amount of information quickly, while low bandwidth channels convey information at a slower rate.

Co-located crowds have a large number of high bandwidth channels of different media types due to the richness of the shared media space. Dispersed crowds may also have these same channels, but at a lower bandwidth. Alternatively, they might only have a subset of these channels, and some channels might require mediation by technology.

4 Categorisation of existing research

With the lens of the taxonomy in place, existing work in the crowd computer interaction domain can be categorised and compared. This section will place existing work into each of the identified dimensions of the taxonomy with justification.

4.1 Intentionality

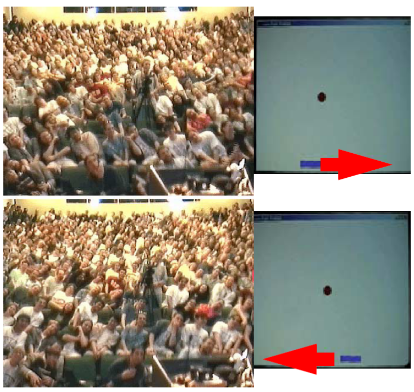

Figure 2: A game of crowd pong is played using crowd leaning as the input to control the paddle (Maynes-Aminzade et al., 2002). This is a classic example of a shared input system: co-operative intentions are created by requiring the crowd to co-ordinate themselves as if a single unit.

Co-operative intentions can be classified into three themes.

First, Aggregate control: the Cinematrix system (Carpenter, 1993) and its derivatives (Fisher et al., 1997; Dannenberg and Fisher, 2001) have applications that aggregate the input of all participants (using the average) and map this to a single control. Applications like crowd pong and collectively steering a car (Carpenter, 1993; Maynes-Aminzade et al., 2002) require the crowd to co-ordinate themselves by turning their paddles to either the red or green side in some required proportion: the system reads the input of all paddles in real-time and moves the paddle or vehicle based on the aggregate of the crowd. The cell membrane game of Fisher et al. (1997), based on the same technology of Cinematrix, requires that the audience balance the ion content of cells by keeping the aggregate value of the paddles at some equilibrium. The same idea is found in the work of Maynes-Aminzade et al. (2002) and derivatives (O2, 2009; Terdiman, 2007) who instead aggregate crowd input using machine vision to capture crowd leaning (see figure 2).

Second, Shared control: requires the crowd co-ordinate a single control around crowd members. The games of Missile command (Maynes-Aminzade et al., 2002), Time Bomb (Sieber et al., 2008), and Squidball (Bregler et al., 2005) use large inflatable balls as the input of the system. The spatial property of the crowd is utilised to encourage crowd members to co-ordinate control: members act to move the ball around the crowd to achieve the required input. The Missile Command game uses the shadow of a beach ball as the cursor for the game: crowd members must hit the ball with a certain velocity and direction to successfully destroy the missiles shown on screen. The Time Bomb game divides the audience in two and requires that members closest to the virtual beach ball (shown on a screen that projects a mirror image of the audience) hit it to the opposing side. Squidball requires that the crowd move sixteen helium filled weather balloons to different locations in the crowd to score points. In each of these instances, the crowd needs to work co-operatively to co-ordinate control of the input.

Third, Power-by-numbers: the numbers found in the crowd are utilised to achieve a shared goal. The Red Nose Game (O'Hara et al., 2008) requires that participants push together virtual red blobs shown on a projection of the crowd displayed on a large public screen (see figure 3). Crowds of one city compete with those in others by attempting to put the red nose back together as quickly as possible. Co-operative intentions are created here as more participants join in to complete the shared goal: members work together to bring the pieces back into a whole as quickly as possible. In a similar way the multi-user connect-the-dots game (Maynes-Aminzade et al., 2002) requires that audience members position dots emitted by lasers in the required configurations; members must work together to complete the task. "Connect-the-dots required each audience member to position his laser over a different dot, and since it required the audience to cooperate in order to succeed it was a more social game" (Maynes-Aminzade et al., 2002, pg 5). The traditional Mexican Wave has been augmented with technology to encourage crowd co-operation. The crowd system created by Uplause (Uplause, 2009) requires that members work together to create an effective wave.

In each of these approaches the crowd is encouraged to work co-operatively by requiring the co-ordination of all members to achieve the shared goal. Properties of the crowd such as the number of crowd members and the spatial layout of the crowd are used effectively to encourage the shared we-intentions.

Separate intention is evident primarily in voting systems. Participants all have the same intention during each interaction: vote for some preference. Although the preference may differ between crowd members, the intention for interaction is the same (I-intend to vote), as opposed to individuals having different intentions during each interaction. The literature presents variations on the voting theme. The Cinematrix system (Carpenter, 1993) and the work of Maynes-Aminzade et al. (2002) each present methods by which the crowd can vote on selected issues, using paddles and lasers respectively. The cheering system of Barkhuus and Jørgensen (2008) allows crowd members to vote with applause and cheering. The voting metaphor can be extended to other systems like the Hewlett-Packard DJ (HPDJ) system of Graham-Rowe (2001) whereby the DJ system automatically picks songs based on how the crowd reacts to them; a genetic algorithm is used to judge which songs have been successful based on who remains on the dance floor among other metrics. During each interaction with the DJ system, crowd members vote with their reaction to the system. Likewise the dance tempo system of Feldmeier (2003) allows crowd members to vote for the tempo of the song by the amount of activity they engage in during their dancing (a sensor measures the amount of force in the dancing). Although crowd members might have their own intentions during the activity, they share the same I-intention when they interact with the system.

Other systems whereby the participants are presented with a goal, but not required to co-operate to achieve it are also shared intentioned activities. Shared I-intentions can be seen in activities like the Whack-a-Mole and picture reveal games of Maynes-Aminzade et al. (2002): "In a game like Whack-a-Mole, each audience member is involved in the activity for himself, much like a game of soccer played by young children in which everone clusters around the ball" (Maynes-Aminzade et al., 2002, pg 5). The former game requires that crowd members shine a laser on an image of a mole that shows on screen, while the latter has crowd members removing the black layer obscuring a hidden image using their laser pointers.

Individual intention, the final type of intention, is represented by crowd activities with no unifying goal or objective. A pattern seen in all the existing work classified with individual intentions is the open ended input each activity provides.

The collective paint application of Maynes-Aminzade et al. (2002) allows participants to draw with laser pointers on a shared canvas. In this instance, crowd members might create co-operative sub-activities if they engage with one another; however, in the general case without system intervention, participants form their own creative intentions (I-intend to draw such and such). The distinction is made from shared intentioned voting activities as crowd members engage with freedom as to the goal of the action, as opposed to voting systems where the goal is singular (to vote).

The interactive dance club (Ulyate and Bianciardi, 2002) allows participants to influence the music and lighting of the dance club environment by using 'interactive zones' situated around the club. Each zone encourages crowd members to engage with unique interfaces to influence the various parts of the environment. Each individual has their own intentions as to how they will influence the shared environment: much like the canvas in the paint application, the environment around the crowd becomes the canvas and the interaction zones the paintbrush.

The Pirates! game (Falk et al., 2001) is a collective activity very much in the spirit of traditional games like hide and seek: players act as ship captains and move around a shared arena with their own quests and objectives. The pervasive game augments the physical arena with a virtual environment (Falk et al., 2001). Although individuals have their own goals and objectives throughout the game, the activity itself is created by the involvement of all crowd members.

4.2 Interaction setting

Many-to-many interactions, whereby many participants contribute to the output in a given interaction, are the most prolific in the identified existing work. The aggregated input systems of Cinematrix based applications (Carpenter, 1993; Fisher et al., 1997; Dannenberg and Fisher, 2001) and crowd leaning applications (Maynes-Aminzade et al., 2002; O2, 2009; Terdiman, 2007) take input from many members for a given interaction and display this on a shared display. The same is true for others of Maynes-Aminzade et al. (2002): connect-the-dots, whack-a-mole, picture reveal, paint; and the Red Nose Game (O'Hara et al., 2008), which use individual participant input instead of aggregation. The interactive dance club (Ulyate and Bianciardi, 2002) operates in a similar way with the environment (lights, music) as the shared output instead: participants interact with different stations simultaneously, and in many cases share a station with others.

The voting systems require the input of all participants before the output is created during a given interaction: therefore the cheering meter (Barkhuus and Jørgensen, 2008), real-time voting systems (Carpenter, 1993; Maynes-Aminzade et al., 2002), dance tempo (Feldmeier, 2003), and the HPDJ (Graham-Rowe, 2001) all have the many to many interaction style.

The Pirates! game (Falk et al., 2001) requires the input of many participants during each interaction; although it can be played as a single player, the activity is only considered a crowd activity with many participants (which is the intended usage (Falk et al., 2001)). The input of many participants, both the virtual manipulation as well as the physical presence, is taken in within each interaction with the system: the output (game environment as well as the physical environment created by the presence and movement of the players) is then given to all participants.

One-to-many interaction is most represented in co-located applications where crowd members shared a single input mechanism amongst themselves. Single crowd member actions are broadcast to the crowd as a whole during one interaction with the system. The missile command game (Maynes-Aminzade et al., 2002), the time bomb game (Sieber et al., 2008), and Squidball (Bregler et al., 2005) all use an inflated ball (virtual in the case of the time bomb game) to control the application displayed on the large screen; members in the crowd control the application one at a time (or in small isolated groups) while the rest of the crowd spectates. Although the crowd may act as one unit (shared we-intention), the interaction is made by a single crowd member.

Twitter (Twitter, 2010), when used in a real-time capacity, also uses the one to many interaction setting. Participants post their message to the rest of the crowd during a single interaction with the activity; the crowd output (message stream) is manipulated by single individuals during a given interaction with the system.

4.3 Interaction style

Reflective style applications allow participants to reflect on the system state before they interact. Reflection allows strategy, expression, and/or creativity for each interaction with the system.

The Pirates! game (Falk et al., 2001) gives each participant objectives that they must complete to move forward in the game. Players search for treasure, compete in battles, and trade with other players to increase both their rank in the game (measured with experience points), and the amount of money they have for upgrading ships. Participants are able to reflect on the system state during every interaction with the system by observing their individual display and interacting accordingly: there is no pressure to react to a stimuli without consideration of the state. Even in the case of sea battles, whereby one player engages another in proximity, participants are able to reflect on the state of their ship and resources while they engage in the battle, and strategically plan their attack or defence.

The voting applications using the Cinematrix system (Carpenter, 1993), or laser pointers (Maynes-Aminzade et al., 2002) are reflective in that they allow participants to consider and reflect on the options available during each interaction: in the work of Maynes-Aminzade et al. (2002) participants are given a topic to vote on and those with laser pointers in the crowd shine their laser beams on their choice, bringing in prior knowledge from their own experiences as well as the knowledge presented by those yelling in the crowd. In a similar way, the cheering system of Barkhuus and Jørgensen (2008) allows participants to reflect during the performance of the artist that they are to judge, and then bring this knowledge into their interaction with the system.

The remaining applications identified with reflective interfaces use expression as their primary reflective mechanism. The use of Twitter (Twitter, 2010) during real-time events allows participants to reflect on the current thread of conversation on the topic and respond accordingly with their own thoughts and feelings. Again there is no pressure to immediately interact without first reflecting on the crowd output. The interactive dance club of Ulyate and Bianciardi (2002) allows participants to manipulate the music and lighting in the dance environment in a very creative way: participants are free to explore the musical and visual canvas by modifying the various components of the music and lighting. The paint application of Maynes-Aminzade et al. (2002) allows creativity in much the same way: participants are free to express themselves through the images they draw on screen.

Reactive interfaces, in contrast, require participants react quickly with speed and/or co-ordination to the system state. Work classically identified as crowd computer interaction mostly centre around reactive interfaces, with interaction made in the moment. Games like crowd pong (Carpenter, 1993; Maynes-Aminzade et al., 2002); collectively driving a car, connect-the-dots, whack-a-mole, picture reveal (Maynes-Aminzade et al., 2002); and the Mexican wave (Uplause, 2009) require participants react to the output shown on screen (and with the crowd itself for the Mexican wave), and react accordingly for successful input.

The activities that use an inflated ball for control (Bregler et al., 2005; Maynes-Aminzade et al., 2002; Sieber et al., 2008) all require the participant(s) currently engaging with the interface to react to the oncoming ball and steer it in the required direction.

In the Red Nose game (O'Hara et al., 2008) participants are presented with a fragmented red blob that they have to piece back together by pushing the various sections into one. Although participants are able to observe and perhaps reflect on the system state on-screen at any time, the primary interaction mode involves quickly reacting to the position of the largest blob, and moving the closest blob towards it: the simplicity of the task means that the reflection time is minimal at best as other participants will quickly act to move the blob if it is not moved immediately by the player.

The dance tempo system of Feldmeier (2003) allows participants to influence the music and lighting using distributed sensors that detect force: the system measures the amount of activity in the dancing and adjusts the music and lighting accordingly. Participants react to the current state of the music and lighting as they dance, and attempt to influence the intensity up or down with their sensor movement.

The HPDJ system (Graham-Rowe, 2001) involves participants reacting to the music in an indirect way. The HPDJ system measures the response to the various music tracks it attempts in engaging the crowd. Participants are presented with a stimuli (the music) and evaluate this in real-time by reacting with their activity. The dancers don't reflect on the input and judge, as with the other voting systems, as the reactions are mostly involuntary: if the track has a beat that suits a given participant, they will react positively with their dancing; otherwise they might decrease their activity or leave the dance floor altogether.

4.4 Social exposure

Figure 3: The Red Nose game (O'Hara et al., 2008) pictured here has the participants highly socially exposed as their actions are clearly visible and highly associated to the actions they make.

Participants in the Time Bomb game (Sieber et al., 2008) are highly socially exposed as their actions are highly visible, and a clear association can be made between the action and the participant. A large screen in front of the audience displays a mirror image of the crowd; overlaid on top of this is a virtual time bomb (with behaviour like a beach ball) that can be manipulated by the audience using gestures. The audience is divided into two competing teams with the objective to have the bomb explode on the opposing teams side of the tiered seating. During each interaction with the system, a single crowd member becomes highly visible as they attempt to swat the bomb away. The member is able to be readily linked with their action as they are displayed on the large screen near the focal point of the crowd as they hit the bomb. Although the duration of the actual action is short, the effects of the action are potentially longer lasting as members remember fumbles or successful saves. In a similar way participants in the missile command game (Maynes-Aminzade et al., 2002) are also considered highly socially exposed as crowd members hit a real beach ball around the audience to control the action on-screen. This game requires more in the way of precision as the player must hit the ball with both the right force and direction to successfully prevent the missile from hitting the city.

Real-time tweets expose the participants due to the nature of the Twitter service (Twitter, 2010): each tweet is identified by the user who created it and becomes a highly visible part of the event's Twitter stream; the action effects are a highly visible component of the crowd output. The duration of the tweet in the event's Twitter stream can vary based on the number of participants actively engaging at any given time; however, even when the stream is quickly flowing due to a high volume of tweets, crowd members can still choose to focus in on a particular tweet at any given moment as the lifetime of the tweet is much longer than the lifetime of the activity.

The Red Nose game (O'Hara et al., 2008) is considered to have high social exposure due to the configuration of the crowd. The proportion of participants is much lower than the number of spectators as the activity has limitations on the amount of space in front of the screen, as well as the reluctance of others to participate. The exposure plays a big role in this reluctance, with by-standers noting an evaluation apprehension based on a fear that they would be judged by others as they interacted with the system (O'Hara et al., 2008). The low number of participants makes actions highly visible and easily associated with those playing the game.

Like the Time Bomb (Sieber et al., 2008) and missile command games (Maynes-Aminzade et al., 2002), Squidball (Bregler et al., 2005) has participants socially exposed as they hit the helium filled balls. However, as the activity uses sixteen balls that are dispersed throughout the crowd, the focus of crowd members is divergent. The visibility of the action effects is therefore somewhat diminished compared to the other ball based activities.

The interactive dance club (Ulyate and Bianciardi, 2002) has a number of 'interactive zones' that allow small numbers of participants to interact with the system. The small clusters of participants together influence the environment of the dance club. The visibility of the manipulations is moderate as the focus of the crowd is divergent due to the many zones and displays situated around the club. The association of action effects to participants is moderate to high depending on the interaction made and the zone in use: zones like the 'tweak zone' allow only one participant to engage at a time with a high association between them and the effect, while others like the 'infrared zone' track a large group of participants.

Applications with low social exposure make up the bulk of the existing work. Many of the applications have large numbers of participants interacting simultaneously with the effects of a single individual only a minor part of the crowd output. This diminishes the association of the action effects to the participant thus causing only low social exposure. This is true of the Cinematrix system and its derivatives (Carpenter, 1993; Fisher et al., 1997; Dannenberg and Fisher, 2001), the dance tempo system of Feldmeier (2003), HPDJ (Graham-Rowe, 2001), the cheering meter of Barkhuus and Jørgensen (2008), and the Mexican wave (Uplause, 2009). The use of lasers in connect-the-dots, whack-a-mole, picture reveal, paint, and the voting systems of Maynes-Aminzade et al. (2002) almost eliminate the link between participants and the action effects.

The Pirates! game (Falk et al., 2001) has its participants spread throughout the game arena: as such many actions are only visible to those in the direct vicinity of the participant. The visibility of actions is therefore reduced as only small subsets of the crowd see the effects of the participant at a given time.

4.5 Crowd member connectivity

The majority of the existing work is based on the co-location of crowd members. In these settings, the full spectrum of communication channels are available at high bandwidth due to crowd members being in close physical proximity.

The real-time use of Twitter (Twitter, 2010) is the exception to other identified work; it is considered to have low member connectivity as only a single high bandwidth symbolic channel is provided between participants. The utilisation of this bandwidth depends entirely on the ability of participants to structure the content of their messages as succinctly as possible.

5 Discussion

This section discusses the categorisation developed in the previous section, and builds on observations made from existing work.

A trend seen in activities identified as having co-operative intentions is the use of crowd properties to create the co-operative environment. Three mechanisms have been identified from the existing work that utilise these properties in engaging crowds co-operatively:

- **Work-as-one**: crowd members are treated by the system as a single entity. This forces the crowd to work together as if a single person to achieve the shared goal. Examples of this are found in the systems that aggregate crowd input, such as the crowd leaning system of Maynes-Aminzade et al. (2002); and those that use a shared input device, like the beach ball system of Sieber et al. (2008). The success of this mechanism is due to the need to co-ordinate and communicate in order to achieve the required unity.
- **Power-by-numbers**: the activity is made easier or possible as more members participate. Crowd members must pool their collective resources to achieve the shared goal: this requires communication and team based thinking.
- **Division-of-labour**: resources or properties unique to individual crowd members must be exploited by the crowd to achieve the shared goal. In the existing work, the spatial properties of crowd members reflects this mechanism: in the Missile Command game (Maynes-Aminzade et al., 2002), crowd members closest to the ball are required to hit it to prevent the oncoming missiles from reaching the cities. The layout of the crowd is exploited here to fully engage and unify the crowd; at any given moment a participant may be required to participate for the benefit of the crowd.

A second trend seen in much of the existing work is the use of reactive interface styles: the visceral immersion provided by this form of interaction creates activities that are able to engage all crowd members regardless of their background or expertise. One of the primary motivations for simplistic interfaces is the need to avoid crowd segregation based on technology or system knowledge: the requirement to engage many co-located individuals in an activity they have had no previous time to explore demands the learning curve be minor for full crowd involvement. The use of reactive interfaces in audience configurations is entirely appropriate due to the requirement for simple, economical, and engaging activities.

Visceral interaction works well to break social barriers; crowd members are able to become fully immersed in the activity and relate to one another at a low common social denominator, the shared experience. The co-operative intentions seen in the activities with reactive interaction styles suggests this strategy works well to unify the crowd in co-located settings.

However, limiting ourselves to this form of crowd activity alone prevents further exploration of the domain. What might the creativity of a thousand minds working in unison create? What is the consensus of the crowd on that offside play? These questions among others provide ample motivation for exploring reflective interaction styles further.

Of the activities identified with a reflective interaction style in the literature, all have associated with I-intentionality; crowd members reflect on their input with individual goals without regard for the crowd as a whole. Observing the reactive applications identified with co-operative intentions, we see that all of the work relies on crowd members being highly synchronous. In order to allow reflection, each individual would need to be able to observe the view of the output for variable periods of time (reflection time). Reducing the high synchronicity requirement raises issues that should be explored in this space: how will the crowd maintain unity if members are out-of-sync, or individual action times are elongated to support the required reflection? The application of the co-operative mechanisms identified from reactive activities might prove a good starting point for this avenue of enquiry.

Low social exposure is a third trend seen in existing work: crowd members and their actions are kept hidden within the crowd as a whole. This is especially true of shared input activities where each member plays only a small role in the collective input. Other activities like the Red Nose game (O'Hara et al., 2008) and BallBouncer (Sieber et al., 2008) have highly visible actions that are strongly linked to participants which provide strong motivating forces for involvement.

Part of the thrill of engaging in crowd activities is the sense of being a part of something big; in most cases this is simply the knowledge that participation led to some effect in the activity and/or the associated 'kudos' of having been at an event (e.g. attending the World Cup final). In other cases this does not suffice: a criticism of the Cinematrix system (Carpenter, 1993) was that the "approaches bury the individual in a crowd where the person's connection with the action on the screen is marginal at best and held in disbelief at worst" (Dannenberg and Fisher, 2001, pg 3).

Social exposure can act to incentivise crowd members to participate in both positive and negative ways. Crowd activities like BallBouncer (Sieber et al., 2008) encourage participation by accountability: each member is clearly visible on shared display, and so members are held accountable if they don't attempt to hit the virtual beach ball shown on screen. Conversely, crowd members are seen by their team in a positive light when they work hard for the team, or help score a critical point. The Red Nose game (figure 3) provides incentive for participation through the novelty of the activity, as well as the potential embarrassment if called out by the compere (O'Hara et al., 2008). In these instances both positive and negative forces act to incentivise participation and engagement.

Conversely, unwanted social exposure can prevent some crowd members from participating. Care must be taken with activities that have high social exposure to limit embarrassment and allow for differing degrees of participation (Reeves et al., 2010). Giving the choice of social exposure to crowd participants allows the best of both worlds as crowd members are able to choose to bask in the kudos of their achievements, or squash their fumble into obscurity.

Almost all of the identified work bases itself on the traditional definition of crowds, with members sharing a single location: as members share the same media space, the connectivity between each of them is very high. This benefits applications in the literature as the crowd is able to co-ordinate itself well with the rich communication medium provided by co-location. However, the high connectivity also has drawbacks in that the channels between participants are uncontrolled. Reducing the connectivity between crowd members by slowing the bandwidth of channels, or reducing the number of channels, allows a reduction in the synchronicity constraint of the system. This leads to a number of advantages. The reduced load on crowd members allows for richer reflective input systems: there is now time to reflect on the system state and so a greater degree of complexity can be introduced. Twitter (Twitter, 2010) allows reflective symbolic communication to occur while also allowing synchronised activity as observed during 'tweeting' about natural disasters (Sakaki et al., 2010). However, the greatest benefit is the ability for crowd members to determine how much involvement they have in the activity and which part of the crowd they consider themselves a part of. This flexibility fits the heterogeneity of crowd members well (Reeves et al., 2010).

Applications built for virtual crowds have the potential to also offer similar synchronous activity found in co-located settings, while engaging a larger and a potentially more diverse crowd; the facilitation of communication channels over technology solutions is a potential direction of enquiry for crowd applications. Can we create compelling activities that allow members from a diverse range of backgrounds, nationalities, and locations to work together?

From the categorisation of existing work, two application class gaps are apparent:

- **individually intentioned reactive interactions**: this class of application requires that crowd members react to some stimuli with speed and coordination to solve their own goals. This could potentially be achieved by having crowd members react to the shared influence in different ways to meet their goals.
- **shared intentioned reflective interactions**: applications of this class allow crowd members to reflect on their input into a co-operative goal, or act to compete against others with the same shared goal.

The latter application class is a good fit for the virtual crowd, owing to the ability to control the communication medium between participants. Individual views of the output could allow crowd members to reflect on their contribution to some shared activity, or perhaps strategise on their own goals in a competitive one. The move away from co-located settings opens up the potential for richer and more complex interfaces: crowd members have time to reflect and up-skill before they fully participate. Members could be presented with their own personal training programme or sandbox before engaging in the activity with others. The ability to control social exposure could allow for a greater degree of experimentation on the part of crowd participants; an important property for reflective interaction styles as interaction now has purpose, as opposed to the reactive interactions of games like crowd pong (Carpenter, 1993; Maynes-Aminzade et al., 2002), and crowd members are more accountable for their interactions.

The virtual crowd provides the possibility of richer reflective interaction: there is no longer the issue of crowd segregation due to technology as each member is equipped with the required interface for the system, and the synchronicity constraints of the system can be reduced to provide reflection time.

The virtual crowd presents the following requirements for any potential supporting technology infrastructure:

- **The ability to connect a large number of participants simultaneously**: crowd members have to rely on the system to facilitate the connections previously created by virtue of co-location. Eliminating the requirement for co-location increases the potential number of participants; the spatial boundaries of typical crowd settings no longer apply in the virtual domain. In addition, the interface that participants use to interact with the system needs to be pervasive enough that a sizeable crowd can form.
- **Rich communication channels**: the system would need to replace communication channels lost from the move away from co-location. Rich communication channels would be needed to support co-operative intentions as they require that the crowd communicate within itself.
- **Real-time**: any potential system would need to support the necessary synchronicity of the crowd so that all members are under the same influence simultaneously.
- **Economy**: one of the major benefits provided by existing solutions like Cinematrix (Carpenter, 1993) and the machine vision system of Maynes-Aminzade et al. (2002) is the economical way in which they engage the crowd. This benefit needs to be carried over to virtual crowds to justify their creation and support.
- **Distribution**: mechanisms for distributing the crowd application to potential participants are required.

The requirements above dictate a solid infrastructure that is able to meet the demands of large number of connected participants, while remaining economical enough to justify the supported activities. The cloud computing paradigm provides two features that address this requirement: large scale economical computing, and elasticity. A solid infrastructure is attainable through cloud service providers at a very modest cost; one of the biggest benefits of the cloud is the reduced entry barriers it brings to large scale computing (Greengard, 2010). Elasticity allows applications to be scaled up and down rapidly as demand on the application grows or shrinks (Owens, 2010). The utility model of the cloud fits the virtual crowd well as crowds tend to meet only temporarily.

Modern smart phones provide a number of advantages as the interface for crowd based applications: they are pervasive; they contain a myriad of sensors such as accelerometers, microphones, with modern phones also sporting compasses and gyroscopes; they have a large amount of computing power; and they are tightly linked to their owners, providing a private interface. The application marketplaces associated with modern phones provide strong distribution networks for applications on the devices.

The combination of the cloud and modern smart phones has the potential to allow the realisation of co-operative reflective interaction in the crowd computing domain.

6 Conclusion

The traditional definition of crowd is challenged by emerging communication technologies. The use of cloud based technologies such as Twitter (Twitter, 2010) in real-time capacities shows that the co-location of participants is no longer necessary for the collective behaviour commonly associated with the traditional form of crowd. This paper proposes that the crowd be defined by the core components of collective behaviour, synchronicity, shared influence, communication channels, and potential for participation in a shared activity. The activity is proposed as the identity of the collective for the purposes of any crowd based application. The proposed taxonomy presents five dimensions considered important for the categorisation of crowd applications: intentionality, interaction setting, interaction style, social exposure, and crowd member connectivity. Using these dimensions, a gap was identified in the existing work: all of the identified work that encourages shared intentions is based around reactive interfaces, utilising the co-location of crowd members. The virtual crowd motivates the exploration of reflective interactions for co-operative and competitive crowd based applications; participants can be given more reflective time by controlling the communication channels between members. The cloud computing paradigm provides an infrastructure that is able to support the virtual crowd economically. Combining the cloud with sensor-rich modern smart phones provides the necessary ingredients to create compelling reflective activities that allow members from a diverse range of backgrounds, nationalities, and locations to play together.

7 Acknowledgements

This work has been partially funded by Microsoft Research Asia, through project "Crowd and the Cloud: Event Central Social Networking with Mobile Devices".

References

Barkhuus, L. and Jørgensen, T. (2008), Engaging the crowd: studies of audience-performer interaction, *in* 'CHI '08 extended abstracts on Human factors in computing systems', ACM, Florence, Italy, pp. 2925–2930.

Blizzard Entertainment (2010), 'World of warcraft community site'. Accessed 6 August 2010. `http://www.worldofwarcraft.com`.

Bregler, C., Castiglia, C. et al. (2005), Squidball: An experiment in large-scale motion capture and game design, *in* M. Maybury, O. Stock et al., eds, 'Intelligent Technologies for Interactive Entertainment', Vol. 3814 of *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, pp. 23–33.

Carpenter, L. (1993), 'Cinematrix, video imaging method and apparatus for audience participation'. US Patent, Nos. 5210604 (1993), 5365266 (1994).

Dannenberg, R. and Fisher, R. (2001), An audience-interactive multimedia production on the brain, *in* 'Proc. Eighth Biennial Symposium on Arts and Technology', Connecticut College.

Falk, J., Ljungstrand, P. et al. (2001), Pirates: proximity-triggered interaction in a multi-player game, *in* 'CHI '01 extended abstracts on Human factors in computing systems', ACM, New York, NY, USA, pp. 119–120.

Feldmeier, M. C. (2003), Large group musical interaction using disposable wireless motion sensors, PhD thesis, Massachusetts Institute of Technology, Massachusetts.

Fisher, R., Vanouse, P. et al. (1997), Audience interactivity: A case study in three perspectives including remarks about a future production, *in* 'Proc. Sixth Biennial Symposium for Arts and Technology', New London, Connecticut.

Graham-Rowe, D. (2001), 'Don't stop moving', *New Scientist* **172**(2317), 25. `http://www.newscientist.com/article/mg17223173.300-dont-stop-moving.html`.

Greengard, S. (2010), 'Cloud computing and developing nations', *Commun. ACM* **53**(5), 18–20.

Maynes-Aminzade, D., Pausch, R. et al. (2002), Techniques for interactive audience participation, *in* 'Proc. 4th IEEE International Conference on Multimodal Interfaces', IEEE Computer Society, pp. 15–15.

O2 (2009), 'O2 launches worlds first interactive 3d cinema game'. Press release. Retrieved 15 July 2010 from `http://mediacentre.o2.co.uk/content/Detail.aspx?ReleaseID=498&NewsAreaID=2`. Archived at `http://www.webcitation.org/5tYqLZcEL`.

O'Hara, K., Glancy, M. et al. (2008), Understanding collective play in an urban screen game, *in* 'Proc. ACM 2008 conference on Computer supported cooperative work', ACM, San Diego, CA, USA, pp. 67–76.

Owens, D. (2010), 'Securing elasticity in the cloud', *Commun. ACM* **53**(6), 46–51.

Reeves, S., Sherwood, S. et al. (2010), Designing for crowds, *in* 'Proc. 6th Nordic conference on Human-computer interaction'.

Reicher, S. (2002), The psychology of crowd dynamics, *in* M. A. Hogg and R. S. Tindale, eds, 'Blackwell Handbook of Social Psychology: Group Process', Blackwell Publishing, chapter 8.

Sakaki, T., Okazaki, M. et al. (2010), Earthquake shakes twitter users: real-time event detection by social sensors, *in* 'Proc. 19th international conference on world wide web', ACM, New York, NY, USA, pp. 851–860.

Searle, J. R. (1990), Collective intentions and actions, *in* P. Cohen, J. Morgan et al., eds, 'Intentions in communication', The MIT Press, pp. 401–416.

Sieber, J., McCallum, S. et al. (2008), 'Ballbouncer: Interactive games for theater audiences'. Retrieved 16 July 2010. `http://metamanda.com/crowdcomputing/subs/sieber-mccallum-wyvill.pdf`.

Terdiman, D. (2007), 'Game turns moviegoers into human joysticks'. Retrieved 15 July 2010 from `http://news.cnet.com/Game-turns-moviegoers-into-human-joysticks/2100-1026_3-6184662.html`. Archived at `http://www.webcitation.org/5tYttBF0n`.

Turner, R. and Killian, L. (1972), *Collective behavior*, Prentice-Hall, Englewood Cliffs, NJ.

Twitter (2010), 'Twitter'. Accessed 1 August 2010. `http://www.twitter.com`.

Ulyate, R. and Bianciardi, D. (2002), 'The interactive dance club: Avoiding chaos in a multi-participant environment', *Computer Music Journal* **26**(3), 40–49.

Uplause (2009), 'Uplause: Social games for big crowds'. Accessed 15 July 2010. `http://www.uplause.com`.

VandenBos, G. R., ed. (2007), *APA dictionary of psychology*, American Psychological Association, Washington, DC.

Whitworth, B. and Plimmer, B. (1997), 'Towards a general taxonomy of communication settings', *National Advisory Committee on Computing Qualifications 10th Annual Conference* .