

Toward Multiscreen Social TV with Geolocation-Aware Social Sense

online entertainment) and media outlets (including smartphones and tablets). Research from both academia and industry is increasingly working to integrate these features into TV applications.

New paradigms, such as interactive TV¹ and social TV,² have been proposed to provide social and interactive features simultaneously with the TV viewing experience. Interactive TV combines screen media, including on-demand TV program delivery, with value-added services, such as online shopping, banking, and so forth. Social TV integrates social networking, such as Facebook and Twitter, into the TV environment, allowing remote viewers to interact with each other via the TV set.

With the widespread use of portable devices, mobile applications can help further personalize the viewing experience.^{3,4} Recently, consumer video watching behavior has changed dramatically, from one screen to multiscreens, with either sequential or simultaneous usage.⁵ Multiscreen or second screen technology involves parallel companion devices, whereby users can engage in other activities without interrupting the TV programming. Moreover, other solutions attempt to make the TV viewing experience ubiquitous and transferable among different devices to realize TV everywhere.⁶ Virtualized screen solutions have been proposed to achieve synchronized content among distinctive terminals.⁷ In this system, screen rendering occurs in the cloud, and images are then delivered to the client for interactive display.

All these innovations have greatly enriched the traditional TV viewing experience, either by extending interaction or including multiscreen fusion features. However, none of them incorporate traditional TV media with fast-growing geolocation social media, which is becoming the most important way that people get information. To reduce the technology gap, we have designed and implemented a multiscreen, interactive TV system that enhances the watching experience by providing a social response. Specifically, we propose a generic, cloud-centric media platform that integrates Big Data analytics with a set of media storage and processing tools.

Leveraging the proposed platform, we have built interactive TV and social sense systems. The interactive TV system integrates enriched social features, including online text and video chat, video comments, and group watching, and the social sense system leverages the Big

A cloud-based, multiscreen, social TV system can enrich content consumption and the TV viewing experience by incorporating and displaying geolocation-aware social data for users on a second screen.

In recent years, the explosive growth in social networks and portable devices has revolutionized the way information is created, disseminated, and consumed. This innovative evolution posits significant challenges to the conventional information propagation carriers, especially for television. Traditionally, TV only provided consumers with limited programming options. Consumers simply decided whether to receive those programs for specific periods of time. In contrast, social networks together with portable devices offer new media forms (such as user-generated content, instant messaging, and

Data analytics paradigm to mine social data associated with the ongoing program to provide deeper insights. Following that, we designed a multiscreen fusion scheme that can transfer the ongoing program-related social sense to a second screen, using an intuitive human-computer interaction technique.

Our prototype system has been implemented on top of a private cloud at the Nanyang Technological University (NTU) and National University of Singapore (NUS) to demonstrate the concept and evaluate its performance. The feature verification and performance evaluation we present here illustrates the intrinsic social features, deep analytics for media content, and intuitive operations for users.

Data-Driven Social TV Architecture

To provide scalability and cost efficiency in the digital TV value chain, we propose a generic, layered system architecture for our multiscreen social TV system with social sense (see Figure 1). The anatomy of this proposed architecture consists of three fundamental components: a back-end system, a multiscreen social TV system with social sense, and consumer components.

Back-End System

To support upper-layer applications, the back-end system leverages two cloud service models: infrastructure as a service (IaaS) and platform as a service (PaaS):

- In the infrastructure layer, raw information and communications technology (ICT) resources (CPU, storage, bandwidth, and so on) are abstracted into a resource pool and provided in the form of virtual machines (VMs) via virtualization techniques. The capacity of the VMs can be dynamically tailored to adapt to resource demands while maintaining service-level agreements (SLAs).
- In the platform layer, we encapsulate a collection of programming tools on top of the infrastructure. We classify essential media-specific services into four categories: content distribution, data storage, media processing, and Big Data analytics.

Other possible services can also be integrated into this layer. These offerings greatly facilitate the development and deployment of our upper-layer applications.

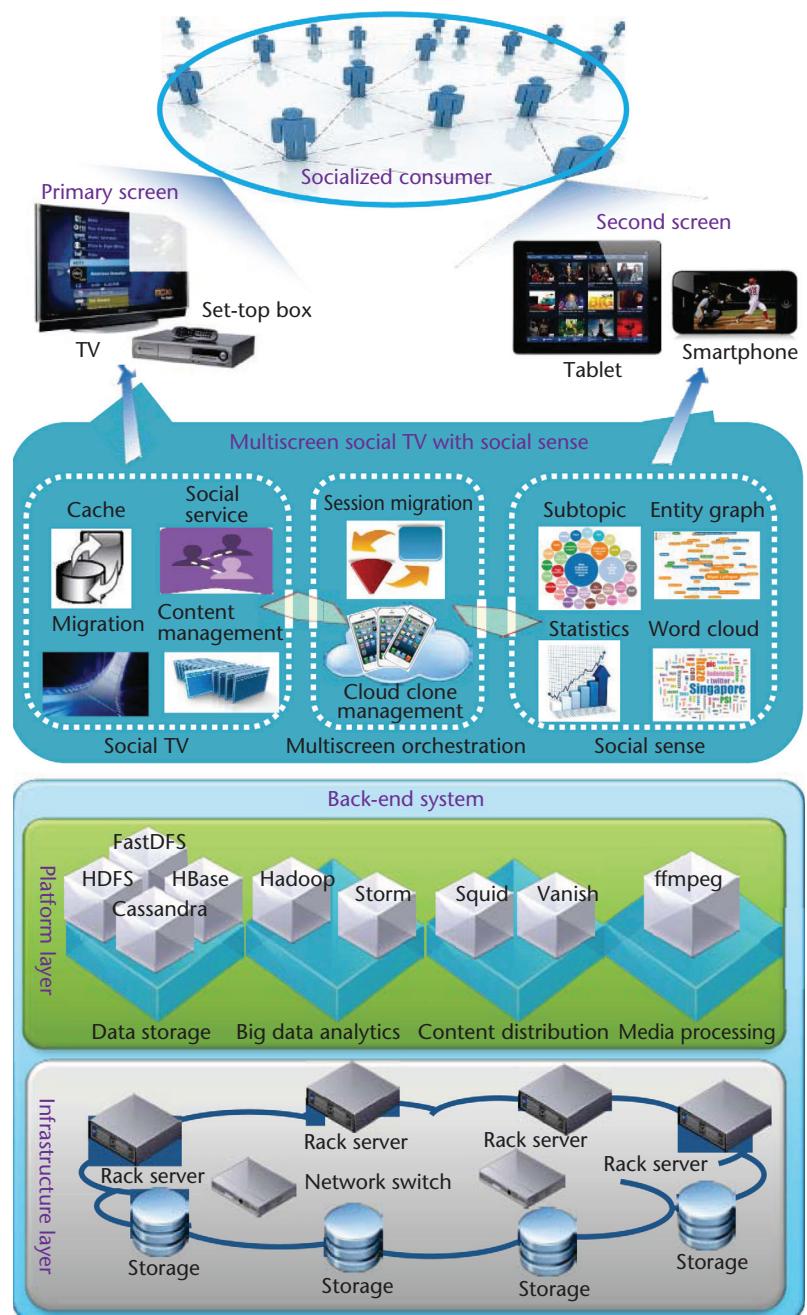


Figure 1. Architecture of the data-driven social TV system. The proposed architecture consists of a back-end system, a multiscreen social TV system with social sense, and consumer components.

Multiscreen Social TV with Social Sense

Our multiscreen TV system with social sense consists of three complementary subsystems:

- The *interactive TV* subsystem implements fundamental TV playback as well as intrinsic social and interactive features. Our system can accept all possible content sources and integrates with popular social

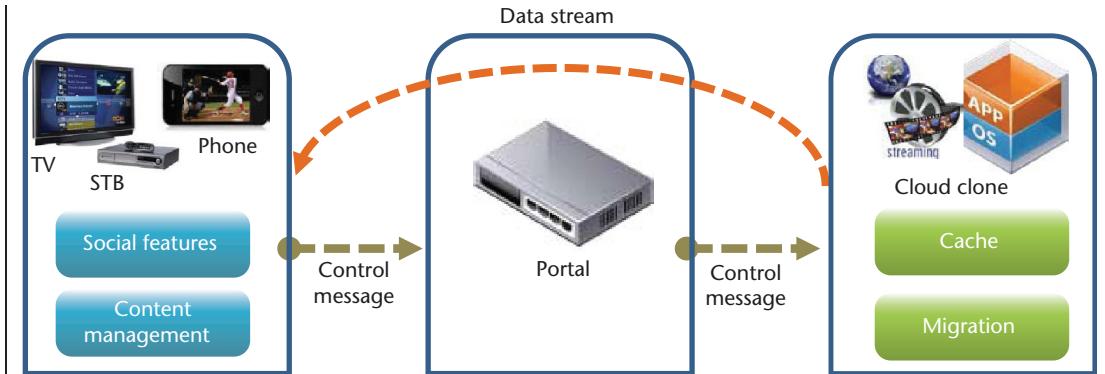


Figure 2. Interactive TV workflow. This subsystem consists of three main segments: media outlet, portal, and cloud clone.

networks, such as Google+ and Facebook, to provide online chat, comment, and group-watching functionalities.

- The *social sense* subsystem crawls social media data associated with TV programs and analyzes them to mine social phenomena, sense, influence, and so forth. Our system can provide profound analysis results, including the extent of social discussion, subtopics, trends, and entity sense.
- *Multiscreen orchestration* incorporates the TV viewing experience with the social sense via a second screen.

Specifically, we display the social sense related to the ongoing TV programming on a second screen with minimal human intervention.

Consumers

By *consumers*, we refer to viewers that use end devices (such as a TV, laptop, tablet, or smartphone) for video consumption. In this work, we consider audiences watching TV programs and browsing the related geosocial observation via a second screen at the same time. However, our system architecture is not limited to any specific media outlet that exists now or will arise in the future.

System Prototype

In this section, we highlight the multiscreen social TV system with social sense in our proposed generic system framework by introducing three key components.

Interactive TV

Figure 2 presents three participatory segments of the interactive TV subsystem: media outlet, portal, and cloud clone. In our system, a cloud

clone refers to the VM that manages all the user's related information, including distinct devices, session, authentication, and so on. We instantiate a cloud clone for each user when he or she requests a session. In particular, a cloud clone serves as the user's proxy, which is responsible of fetching media content, transcoding it for different outlets, and synchronizing status.

A typical TV session consists of three steps:

1. The consumer requests a TV program through a media outlet.
2. The portal accepts the request and creates or migrates a cloud clone for this user.
3. The cloud clone parses the request, fetches the demand media content, and then sends the content to the media outlets via the portal.

To realize this procedure, we customize our media outlet and cloud clone solutions.

Media Outlets. Our system enables audiences to watch TV programs with a TV or any portable device, such as a tablet or smartphone. In particular, we implement two key components: content management and social features. The content management module lets users operate four categories of content resources: local, media sharing center, over-the-top (OTT) online, and cloud. The social feature module integrates our system with Facebook and implements an internal instant messaging mechanism for text and video chatting and commenting on ongoing videos.

Cloud Clones. Cloud clones can help serve large-scale audiences from geolocation distributed areas with varying network environments.

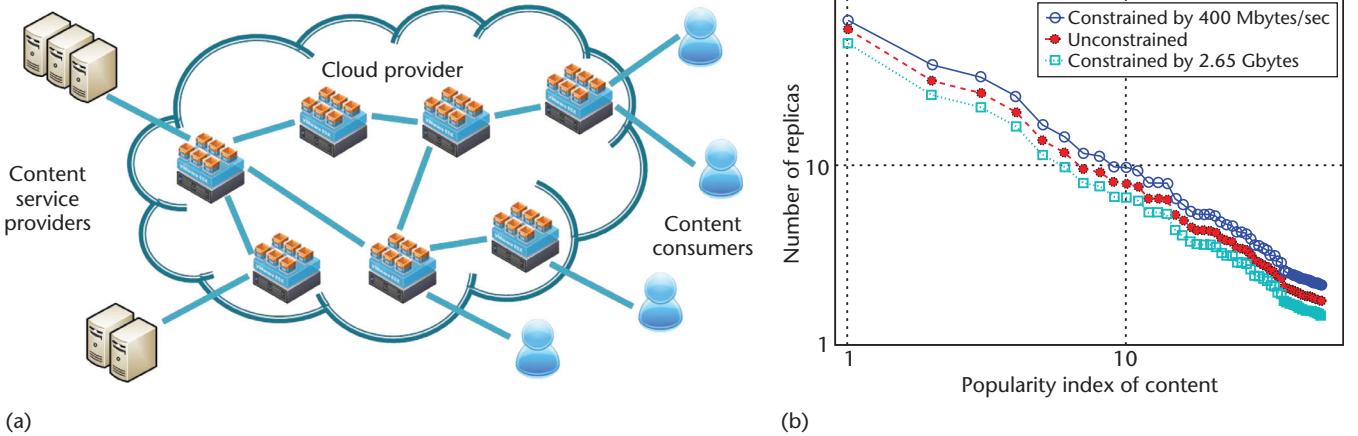


Figure 3. Geolocation-aware content caching. (a) System architecture and (b) optimal number of content replicas.

There are two obstacles to achieving this goal. One is to distribute the rapidly growing amount of user-generated content (UGCs). This is a challenge because the request population, content popularity, and resource capacity are all geolocation diverse. The second challenge is that users with different media outlets and distinct bandwidth requirements exhibit diverse watching behavior. Thus, cloud clones must migrate along the transmission path to save bandwidth consumption. We tackle these two problems by using geoaware content caches and geobased cloud clone migration.

Content caching distributes media content into different cloud edge servers to reduce the distance between users and content, leading to an improved end-user experience. Figure 3a illustrates the system architecture. However, determining the number of replicas and their locations is essential to managing operational costs. This involves a fundamental trade-off. On the one hand, more replicas can reduce the distance between the users and content, which can translate into networking cost. On the other hand, if too many copies are created and their locations are not chosen properly, significant storage cost can incur with limited gains in networking costs.

We can formulate this problem as a constrained graph optimization problem⁸ as follows: Let's assume a topology $G = (V, E)$, where V is the set of edge servers and E is the set of links between them. The goal of content caching is transformed into computing a subset of n vertices that minimize the mean hop distance between any two nodes for a given amount of content. The numerical verifications (see Figure 3b) in the US IP backbone network

suggest that the optimal number of replicas follows a power-law distribution with respect to its popularity.

In our multiscreen system, cloud clones fetch content from the aforementioned caching system, transcode it to the required bit rate, and transfer it to media outlets. The cloud clone's location is within the transmission path, as Figure 4a shows. The media content's original bit rate of is B_0 , whereas the transcoded bit rate for the media outlet is B_t , which is determined by the type of active end device. When $B_0 < B_t$, placing the cloud clone at the node nearest the user can minimize the transmission cost; otherwise, the farthest node is preferred.

However, cloud clone migration incurs extra transmission overhead. The challenge is to find an optimal scheduling policy over time to balance the transmission and migration costs. We can formulate the problem as follows:⁹

$$\min \sum_{t=0}^{T-1} (C_{\text{tr}}(l_k(t)) + C_{\text{mig}}(l_k(t+1), l_k(t)))$$

where $l_k(t)$ is the location of the cloud clone at time t ; $C_{\text{tr}}(l_k(t))$ is the transmission cost; and $C_{\text{mig}}(l_k(t+1), l_k(t))$ is the migration cost determined by the location distance at successive times. We use a Markov decision process (MDP) to solve the user-behavior-driven optimal location strategy. The result, as illustrated in Figure 4b, suggests that the cloud clone can only be located at the nodes nearest to or farthest from the user, regardless of the delivery path length.

Social Sense

Microblog services provide a platform for users to share everyday thoughts, opinions, and

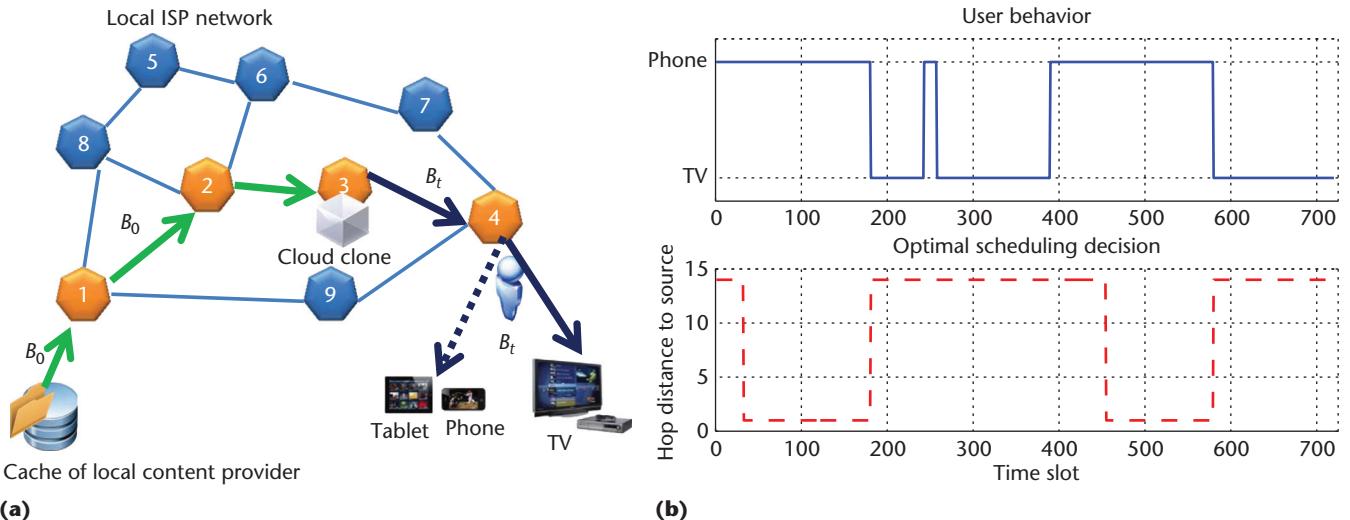


Figure 4. User-behavior-driven cloud clone migration. (a) Cloud clone migration and (b) optimal migration policy.

experiences. Parts of this UGC reflect and reveal their interests, concerns, and criticisms about TV programs. The aim of geolocation social sense is to associate the public perception with ongoing TV programs. The system consists of two stages: relevant data collection and emerging characteristic modeling of detected topics.¹⁰

To mine the social sense from UGC, the first step is to crawl a relatively complete set of messages associated with the designated media content. This is not a trivial task because most of the live microblog services set limits on the amount and frequency of data that can be crawled. In addition, because of the size limit on microblog messages, many related messages do not contain the expected keywords, so the relevant ratio to specific media content is usually low.

To tackle this problem, we designed four types of crawlers: fixed keywords, dynamic keywords, known accounts, and key users. First, fixed keywords are manually selected to uniquely identify the media content. Similarly, known accounts are manually selected to identify a set of media-content-related users, such as an ongoing program's official account or that of its director. Dynamic keywords and key users are then extracted from the tweet sets crawled by the fix keywords and known accounts.

The second step uses machine learning, text, and image analytics techniques^{11–13} to discover knowledge from the data, such as media context, geolocation, and key users. An online or incremental clustering algorithm is first used to

discover topics that guarantee real-time performance. Then, we analyze the emerging topic-related features, including user authority, tweets influence, and emerging keywords. These features are incorporated into a topic learner to identify the emerging topics in a timely manner. The social sense is determined by providing the analysis results in terms of the statistics of crawling data, emerging topics, key entities (organization, person, location, and miscellaneous), and trending words.

Multiscreen Orchestration

The general idea behind multiscreen orchestration is to build a link between TV programs and their social sense. This feature allows consumers to scan the quick response (QR) code on the TV screen using the camera on their second screens to obtain the related social sense. Two main components are involved: cloud clone management and session migration.

Because each user corresponds to a unique cloud clone, the profiles and status of all the applications running on media outlets can be orchestrated by a cloud clone via an interdevice message bus. As a result, we can easily search and synchronize the status of distinct devices belonging to the same user. To achieve fast routing and information retrieval, all the cloud clones form a logical ring via a distributed hash table (DHT). In the DHT key space, each cloud clone is uniquely determined by a key. As a result, the route length can be limited to $O(\log n)$, where n is the total number of nodes in the DHT ring.

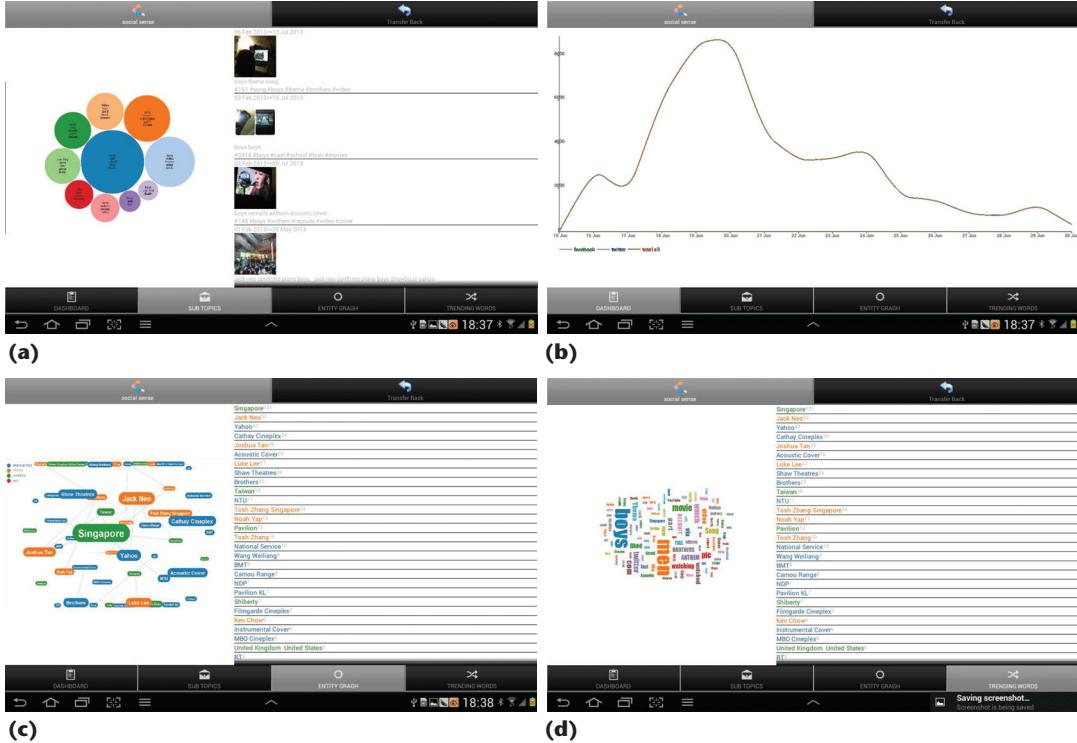


Figure 5. Social sense features. (a) Topics, (b) degree of interest, (c) entity graph, and (d) keyword cloud.

To start a session migration, users must log in to the cloud to get authenticated. Upon confirmation, users can capture a QR code on the TV screen and send the session migration request to the cloud. Based on this request, the cloud clone will recognize the session and make confirmations accordingly. After the users receive this confirmation message, they can eventually trigger the session migration and obtain the social sense data.

This scheme has several advantages. First, it supports session migration without requiring that users recall and type a password. Users only need to register all their terminals and bind them together in the beginning. Second, users often have trouble understanding new applications. Our scheme offers users a simple and intuitive procedure that should help to improve their learning curve in practice. Finally, our scheme provides a generic link between TV and mobile terminals that is compatible with other applications.

Feature Verification and Performance Evaluation

In this section, we first demonstrate the salient features of our system. Following that, a subjective evaluation illustrates the operational

effectiveness of our multiscreen orchestration scenario.

Testbed Implementation

We build our system on top of a modular data center at NTU that consists of 10 racks. Each rack contains up to 30 Hewlett-Packard servers and 2-Gbit Cisco switches. The data center can provide an ICT capacity of a 25-Tbyte disk, 1,200-Gbyte memory, and 600 CPUs. We used CloudStack on CentOS 6.3 as the host operating system to construct the infrastructure layer, which virtualizes physical servers into a collection of VMs. Based on our infrastructure layer, we encapsulated a set of data storage, indexing, and processing tools, including the Hadoop suite, FastDFS, Cassandra, SOLR, Storm, and ffmpeg, to form the platform layer. The social sense platform is located at NUS.

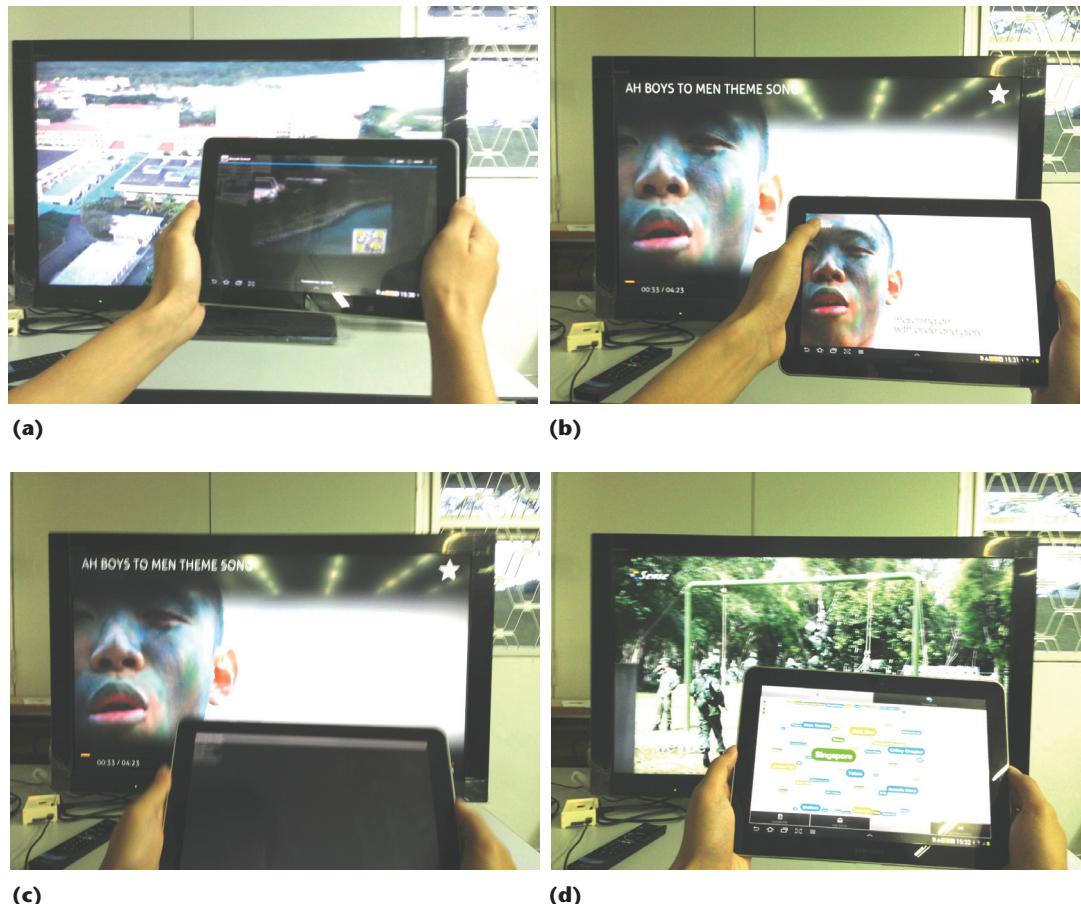
Feature Verification

Our system provides four aspects of social sense, including topics, degree of interest, entity graph, and word cloud (see Figure 5). The degree of interest refers to the potential interest and popularity of the subject or topics, as reflected by the message count. From Figure 5a

Figure 6. Social sense of a famous Singapore local film, Ah Boys to Men. (a) Tweets related to the topic “time Singapore film” and (b) entities related to the film, including people, organizations, location, and miscellaneous.



Figure 7. Steps to complete a session migration. (a) Screen shot, (b) interest indication, (c) gesture triggering, and (d) social sense acquisition.



and Figure 5b, we can identify emerging topics and the corresponding degree of interest. The entity graph (see Figure 5c) shows who is talking about the media content in terms of the person, organization, location, and miscellaneous. The keyword cloud (see Figure 5d) shows the set of keywords extracted from crawled messages and their weight.

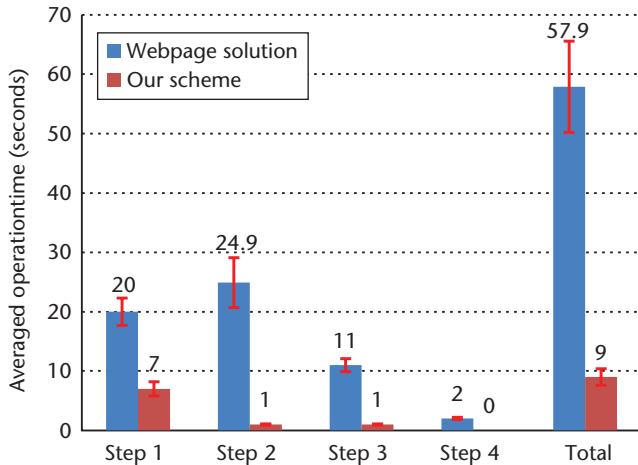
Figure 6 shows an example from a famous Singapore film, *Ah Boys to Men*. Users can see all the topics about this film from Figure 5a, including "Song theme my brother," "time

Singapore film," and so forth. They can click one of the topics in the topic cloud to see all the related messages (see Figure 6a). Moreover, all the relevant people, organizations, and locations about this film are graphed in Figure 6b, which can be accessed through the link in the entity graph. (For more information, see a video demo on YouTube at www.youtube.com/watch?v=UQBcBoUGEbg.)

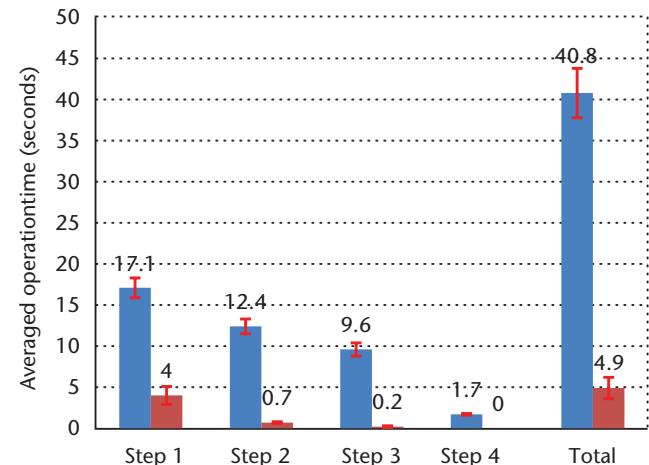
Figure 7 shows the intuitive multiscreen orchestration process, which consists of four steps. First, users hold their mobile phone or

Table 1. Operational procedure comparison for a traditional webpage solution and our scheme.

Step	Webpage solution	Our scheme
1	Go to the webpage	Screen shot
2	Log in	Interest indication
3	Search the content	Gesture triggering
4	Choose the social sense	Social sense transmission



(a)



(b)

Figure 8. Averaged operation time. (a) First and (b) 10th times.

tablet in front of the TV to obtain the authentication details through the QR code (see Figure 7a). They then press the sense icon to select the social sense information (see Figure 7b). Once the cloud is ready for social sense transmission, users can trigger the process by performing a flip gesture (see Figure 7c). After those simple steps, the social sense information is displayed on the mobile phone or tablet (see Figure 7d).

Subjective Evaluation

We conducted a user study to compare our scheme with a traditional webpage solution in terms of retrieval time. Less retrieval time indicates a smoother operational procedure resulting in a better user experience, a shorter learning curve, and better operational success ratio.

Using a traditional webpage, participants access the social sense webpage, log in to an account using a password, search the media content, and choose the expected social sense information. Table 1 compares this procedure with our solution.

To time our operational procedure, we recruited 15 participants from NTU students with majors in computer science. Seven of the participants were female, and eight were male. During the experiment, we first asked participants to acquire social sense information using the webpage solution and recorded the operation time for each step separately. We then requested that they repeat this procedure 10 times. Following that, the same flow was conducted for our multiscreen orchestration scheme.

Figure 8 shows the average retrieval times and their standard variations for the two solutions for the first and 10th access times. In the beginning, all the participants were unacquainted with the two solutions. We compared the operation times to show the users' learning curves. As Figure 8a shows, the first total operational time was 57.9 seconds for the webpage solution and 9 seconds for our scheme. This indicates that our scheme is more user friendly, allowing a faster learning curve.

After 10 times, all the participants were familiar with the operation procedure. The final

Our scheme outperforms the webpage solution with a shorter average operation time.

operation time was 40.8 seconds for the webpage solution and 4.9 seconds for our scheme, as Figure 8b shows.

Clearly, our scheme outperforms the webpage solution with a shorter average operation time. In particular, we can see that the majority of the time was consumed during steps 1 and 2 for both solutions. However, our scheme uses QR code scanning to replace the typing keyboard operations, resulting in less operational time.

Conclusion

We are currently building a unified Big Data platform for social TV analytics. Such a platform integrates several components to facilitate social media analysis, including a social media crawler, Hadoop, Storm, and other natural language processing tools. It aims to benefit stakeholders in the traditional TV ecosystem. **MM**

Acknowledgments

This research was partially supported by Singapore MOE tier-1 fund (RG 11/31) and NTU SUG, the Singapore National Research Foundation under its International Research Center @ Singapore Funding Initiative, and administered by the IDM Programme Office; the National Basic Research Program of China's 973 Program (grant 2012CB316400); the National Natural Science Foundation of China (grant 61125106); and Shaanxi Key Innovation Team of Science and Technology (grant 2012KCT-04).

References

- M. Ursu et al., "Interactive TV Narratives: Opportunities, Progress, and Challenges," *ACM Trans. Multimedia Computing, Comm., and Applications*, vol. 4, no. 4, 2008, pp. 25:1–25:39.
- P. Cesar and D. Geerts, "Past, Present, and Future of Social TV: A Categorization," *Proc. 2011 IEEE Consumer Comm. and Networking Conf. (CCNC)*, 2011, pp. 347–351.
- R. Ji et al., "Location Discriminative Vocabulary Coding for Mobile Landmark Search," *Int'l J. Computer Vision*, vol. 96, no. 3, 2012, pp. 290–314.
- Y. Gao et al., "W2go: A Travel Guidance System by Automatic Landmark Ranking," *Proc. ACM Int'l Conf. Multimedia*, 2010, pp. 123–132.
- Google, Ipsos, and Sterling, "The New Multi-screen World: Understanding Cross-Platform Consumer Behavior," Aug. 2012; http://services.google.com/fh/files/misc/multiscreenworld_final.pdf.
- R. Martin et al., "neXtream: A Multi-device, Social Approach to Video Content Consumption," *Proc. 7th IEEE Consumer Comm. and Networking Conf.*, 2010, pp. 1–5.
- Y. Lu, S. Li, and H. Shen, "Virtualized Screen: A Third Element for Cloud Mobile Convergence," *IEEE Multimedia*, vol. 18, no. 2, 2011, pp. 4–11.
- Y. Jin et al., "Toward Monetary Cost Effective Content Placement in Cloud centric Media Network," *Proc. Int'l Conf. Multimedia and Expo (ICME)*, 2013, 2013, pp. 1–6.
- Y. Jin, Y. Wen, and H. Hu, "Minimizing Monetary Cost via Cloud Clone Migration in Multi-screen Cloud Social TV System," *Proc. Int'l IEEE Global Comm. Conf. (Globecom)*, 2013, pp. 1747–1752.
- Y. Chen et al., "Emerging Topic Detection for Organizations from Microblogs," *Proc. 36th Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*, 2013, pp. 43–52.
- Y. Gao et al., "Visual Textual Joint Relevance Learning for Tag-Based Social Image Search," *IEEE Trans. Image Processing*, vol. 22, no. 1, 2013, pp. 363–376.
- R. Ji et al., "Task-Dependent Visual Codebook Compression," *IEEE Trans. Image Processing*, vol. 21, no. 4, 2012, pp. 2282–2293.
- R. Ji et al., "Learning to Distribute Vocabulary Indexing for Scalable Visual Search," *IEEE Trans. Multimedia*, vol. 15, no. 1, 2013, pp. 153–166.

Han Hu is a research fellow with the School of Computing at the National University of Singapore. His research interests include social media distribution and analysis. Hu has a PhD in control theory and control engineering from the Department of Automation at the University of Science and Technology of China, Hefei, China. Contact him at huh@comp.nus.edu.sg.

Yonggang Wen is an assistant professor in the School of Computer Engineering at the Nanyang Technological University (NTU), Singapore. His

research interests include cloud computing, content networking, and green networks. Wen has a PhD in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT). Contact him at ygwen@ntu.edu.sg.

Huanbo Luan is a senior research fellow with the School of Computing at the National University of Singapore. His research interests include multimedia analysis, mining, and retrieval. Luan has a PhD in computer science and technology from the Institute of Computing Technology, Chinese Academy of Sciences. Contact him at luanhb@comp.nus.edu.sg.

Tat-Seng Chua is a full professor with the School of Computing at the National University of Singapore. His research interests include multimedia information processing, including the extraction, retrieval, and questioning and answering of text and video information. Chua has a PhD in computer science

and technology from the University of Leeds, UK. Contact him at chuats@comp.nus.edu.sg.

Xuelong Li (corresponding author) is a full professor with the Center for Optical Imagery Analysis and Learning, State Key Laboratory of Transient Optics and Photonics, Xi'an Institute of Optics and Precision Mechanics at the Chinese Academy of Sciences. His research interests include visual surveillance, biometrics, data mining, multimedia information retrieval, and industrial applications. Li has a PhD in signal and information processing from the University of Science and Technology of China. He is a fellow of IEEE. Contact him at xuelong_li@opt.ac.cn.

CN Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.

NEW IEEE computer society STORE

Find the latest trends
and insights for your

- presentations
- research
- events

webstore.computer.org

Save up to
40%

on selected articles, books,
and webinars.

