iRain: Data Water Cycle for Raining User with Information

Ganesh Venkatesh Sravanthi Kota Venkata Robert Chen
Steven Swanson Michael Bedford Taylor

The market for smartphones and portable devices has been growing rapidly making it possible to remain connected all the time. This trend is only expected to continue as more portable devices such as iPad, HP Slate and faster smartphones such as Nexus One hit the market. Meanwhile, storage is becoming cheaper leading to rapidly increasing laptop and desktop storage, which is often under-utilized. On the other hand, portable battery driven devices tend to have limited disk storage because of power and weight limitations. In addition, previous research projects such as OceanStore focused on using internet as *Data Ocean* for storing the entire user content across it.

In this project, we propose the data "water cycle", iRain, that ensures that the user is "soaked" with useful user data all the time. Figure 1 shows the iRain system architecture consisting of *iCloud server, iRain client, iBrella Data Filter, and user devices (iPuddles)*. While the file system interface provided to the user remains the same, the OS would include the local iRain client. The iRain client monitors user activity to determine the important files. These might include the recently modified text or media files, result files for the experiments user ran, frequently accessed files etc. The iRain would "evaporate" this data to the iCloud server. One of the challenges for iRain system would be to ensure no *iThunderstorms* are caused by evaporating too much data. For instance, if the user makes wholesale changes by modifying few strings in almost every file on the disk, then iRain should not evaporate all of them even though they were recently modified.

The iCloud server seeks to maintain a cloud of user data around the user by storing it on the iPuddles present in the proximity of the user. This would ensure that users can access their content on the portable devices while travelling without having to explicitly upload or download anything. This distributed technique ensures minimal bandwidth requirements for iClient server because much of the content is accessed by peer-to-peer transfer. Also, the peer-to-peer structure enables parallel transfer of the content to the user device to improve the end-user experience.

The iCloud server would routinely "rain" the user's portable devices with data. The portable devices have iBrella for shielding from unnecessary incoming data decided based on the usage characteristics of the device. For instance, iCloud might rain the iPhone with recently modified video file along with intermediate versions and other related information needed to make further changes. The iBrella on the iPhone might only accept the final modified video and ignore the rest because of the limitations in space and video editing

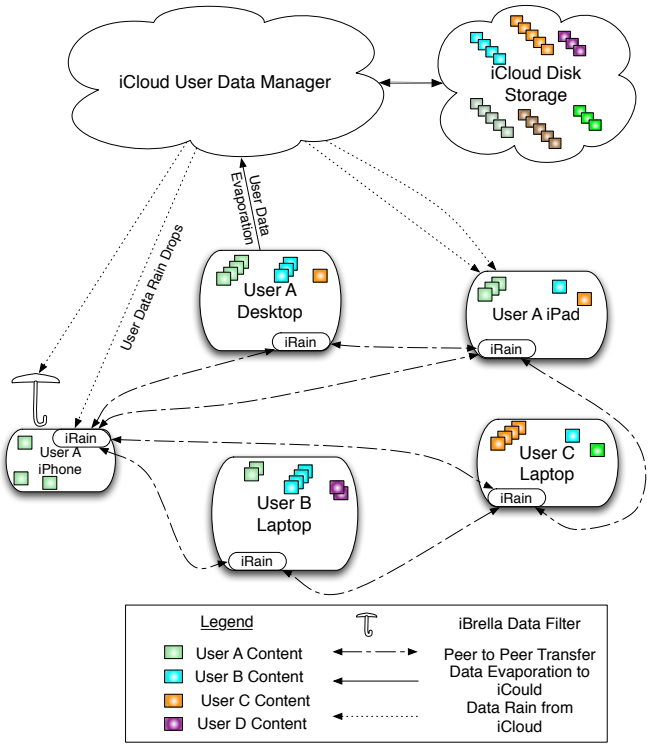

Figure 1: **iRain System Architecture**. The iRain System Architecture is shown in this figure. The user data deemed "useful" by iRain evaporates to the iCloud server. The iCloud rains the user's devices with the data as well as maintains a cloud of user data in proximity to the user. The portable devices have iBrella data filter to only soak in the data considered useful for that device.

capabilities. Similarly, iBrella might ignore log files and only accept the final result table for the experiments recently run by the user.

In conclusion, the work presented here aims to enable users to access their information anywhere on any device seamlessly without any manual synchronization. Many of the components used in our system already exist; such as CHORD, OceanStore, Bit-Torrent focus on storing and accessing data across internet and peer-to-peer architectures. Our work maximizes utilization of the limited storage available on the portable devices by only storing important information on them. The success of our system would ultimately depend on the accuracy of iRain and iBrella in identifying the useful user data and making it available to the end-user in a non-intrusive, timely, and power-efficient manner.