Candace Ming

Prof. Besser

H72.1807

12/15/10

MLB Archive

The MLB archive, located in Secaucus, NJ, is a sprawling facility that houses the production, network, and archive arms of Major League Baseball. The facility has been operational for just two years after the launching of the MLB network in 2008. The infrastructure was created to provide interoperability between the archive, production, and the network. The archive ingests material from both the production arm as well as the network. The archive is also in the process of digitizing their analog holdings using a digital system that was designed by a company called Grass Valley. Grass Valley organized their storage systems as well as facilitating communication between departments. The archive also created DIAMOND (Digitized Industry Assets Managed Optimally for Networked Distribution), a media asset management system that logs, tracks, sorts and facilitates mining and retrieval of material. The building itself is laid out in a somewhat open format, with each department having distinct and separate spaces. For the purposes of this report I have broken down each department into sections to communicate more effectively their digital infrastructure and management.

MLB Legacy Archive

The archive is the repository for all of the live game feeds, metadata, and production materials. MLB has moved to a digital file based infrastructure, but the archive has legacy materials in the form of BetacamSX tapes. The archive migrated it's

2”, 1”, and ¾” holdings over the last ten years to the BetacamSX format. The archive has footage dating back to the early 1900’s all the way up the 1990’s. The archive uses the SAMMAsolo and SAMMAclean in their digitization workflow. The archive also has a dedicated SAN known as SAN-W. Tapes are cleaned using the SAMMAclean machine and then placed in the SAMMAsolo, which creates four different types of digital files. A JPEG 2000 preservation master is created, along with a Windows Media File, H.264 file, as well as an HD file that can be used by production. These files are then moved to a tape library, which creates two copies, one for on-site storage and one for off-site. The Grass Valley structure wraps each file in its own format known as a GXF file. This allows MLB to point to individual files in a package (such as an individual audio or video track), is independent of any codec, and allows for 16 tracks of audio. The GXF format also supports metadata. Using a program called Aurora the archive can view, browse, and sort their content on the desktop. Grass Valley also creates a linking system for editing and playback of files. Grass Valley keeps the actual clip on one server and “winks” at it for playback purposes. This reduces the storage needed and anything done to the virtual clip is saved to the actual clip. Currently the archive has over 150,000 hours of footage.

The tape library has over 3,000 slots and serves as the final repository for production and archive materials. The archive currently uses LTO-4 tape, but plans to migrate when LTO-7 is released. Two copies are made of any program for storage on and off-site. Aurora interfaces with the tape library software making it easy to retrieve tapes for editing or playback purposes. The tape library delivers contents at a speed of 300-400 Mb/s, which delivers content within minutes of a request. The tape library is also cost-efficient. Each tape holds 35 hours as opposed to the BetacamSX tapes, which are only

184 minutes. The tape library is also used because of the massive storage requirements of the MLB network. Around 375 hours of unique content are created per day at the network and in just 2009 MLB utilized 2,500 tapes. MLB estimates that they will use between 9,000-12,000 tapes this year alone.

MLB Production Archive

MLB productions records each game played on a given day with sometimes as many as 15 games being recorded or played each day. Home and Away games are recorded and transferred to the archive through OC3 fiber optic cables. The games are encoded in XDCAM HD50 and transmitted at 50 Mb/s. Live feeds stream at 950 Mb/s for faster retrieval and editing purposes. Editors have the ability to edit live content and push it to the network with only a three second delay. The production archive receives a "clean" version of each game (without graphics) for editing. Each league has their own dedicated SAN, SAN-X (See Index 1) for American League and SAN-Y (See Index 2) for National League games. Each SAN holds around 4,000 hours of footage. Programs like Aurora and Final Cut Pro as well as each computers hard drive take about half of the storage capacity of each SAN. A quick turnaround for editing games is necessary so that new content can come in each day. Another SAN-Z is used to park material until it can be moved to the tape library. SAN-Z can hold up to 18,000 hours of footage. Each stadium also has a remote controlled ballpark cam feed that goes directly to the Production archive. Production has 18 Aurora and 15 Final Cut Pro edit suites. Aurora is mainly used for quick turn-around of games, while Final Cut Pro is utilized for editing longer

programming. The GXF wrapper allows all of these systems to integrate since each program recognizes different file formats. Final edit versions can be sent to Omneon, the on-air system for use in programming.

A data feed is also transmitted from each ballpark and ingested into the DIAMOND database. The feed includes game play statistics, which are transmitted to logging stations at the archive. Loggers can tag plays as “highlight worthy” as well as identify names. This data is re-imported to DIAMOND and is often used by editors compiling highlight reels. Loggers are able to pause live games and rewind in order to tag certain plays. A touch screen system is used with categories that can be pressed like “triple play” or “pop fly.” This data can be exported by XML back to Grass Valley for use in editing clips. DIAMOND also uses this data to populate the ticker and scoreboard on-air.

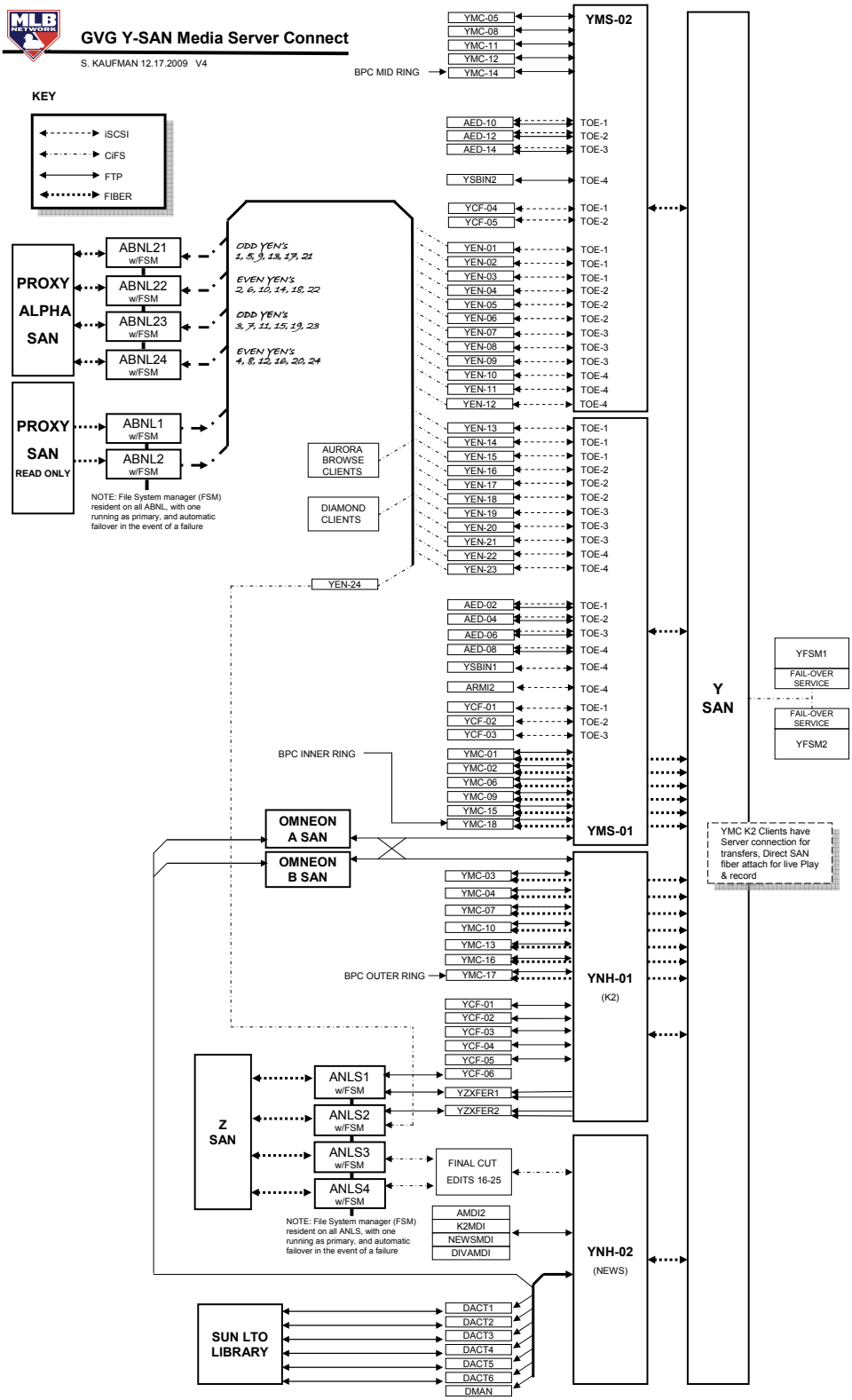
MLB Network

The network uses the Omneon system to retrieve and access content. The network also has two dedicated SANs to store content pushed from production or the archive. During live games the network directly receives a dirty version (with graphics) of every game being played on a given day. The GXF wrapper allows the network to broadcast in XDCAM HD50 with 16 tracks of audio. Omneon rewrote their interpreter to recognize this file type because the standard XDCAM HD50 only allows for 8 tracks of audio. At the Secaucus facility there are two studios the network uses. One is a basic set and the other is model ballpark for recreating plays and memorable moments. There is also a control room and monitoring station. The network utilizes a system called Reality Check

Uppercut, which allows a producer and operator to remotely operate a camera as well as program. The system is mainly used for reporting breaking news and programming lower thirds.

The entire infrastructure of the MLB facility was created to maximize efficiency and turn around time. In a typical day, MLB network is live 12-14 hours and streaming two different feeds (a clean and dirty version) to different parts of the archive. The ability for editors to access and push content quickly is key to the archives structure. Storage is also a large part of the archive with each year increasing the content and materials needed to be stored. Currently, the archive is looking to upgrade their Grass Valley software and is monitoring other companies that use the same infrastructure.

Index 1: Wiring Diagram for SAN-Y



Index 2: Wiring Diagram for SAN-X