GridCast
A Service Architecture for the Broadcasting Media

Terence Harmer, Christina Cunningham, Paul Donachy, Julie McCabe, Mark McColgan, Ron Perrott and Gareth Waller
Belfast eScience Centre
The Queen's University of Belfast
Belfast BT71NN

Stephen Craig, Chris Chambers, Luke Sluman, Brendan Mallon and Kevin Price
British Broadcasting Corporation
Broadcasting House
London

Abstract

The **GridCast** project is piloting grid services to support the BBC Nations in their broadcasting activities. The project has developed datagrid-like services together with processing services that support wide-area data distribution and processing. The challenges in developing grid services for broadcasters are the high reliability, resilience, security and reactivity requirements that broadcasting requires. The project has developed a prototype media grid infrastructure to act as a test bed for media services.

1. A Broadcast Media Grid

The GridCast project is a collaborative industrial eScience project that aims to develop a prototype media grid that provides a media content sharing and media services environment for the BBC. The focus of the project is the support of BBC National broadcasters, such as BBC Northern Ireland, BBC Scotland and BBC Wales (called the BBC Nations), in their provision of tailored broadcasting schedules for their communities. (However, the project is acting as a test bed for ideas for a larger-scale broadcasting, archiving and programme production grid for the BBC.)

The BBC Nations are increasingly defining their own broadcast content and controlling more of their broadcast schedule. This trend will require the implementation and management of a distributed broadcast content sharing network along with the necessary media support services to process content before it is broadcast. In the current BBC network infrastructure, broadcast content is stored centrally in London on media servers and distributed, via leased ATM circuits, to the BBC Nations as if it were a live programme. Thus, for example, an EastEnders episode will be taken from a media server in London and be distributed to the BBC Nations in the same fashion as a live News broadcast. In the GridCast broadcasting model, recorded content will be copied to the BBC Nations when the Nations request the content and when it becomes available, and a network-wide content management system implemented. The aim is to provide a broadcast infrastructure that is flexible and cost-effective.

2. The Issues in Broadcast Media Grid

Broadcasting is a fast moving, highly competitive and highly reactive industry. It is an industry that is moving from one that was dominated by special-purpose tools and infrastructure to one that uses a conventional IT infrastructure. Thus, an industry that was dominated by special-purpose editing facilities increasingly uses desk-top editing tools, where broadcast content was once stored on tape is increasingly stored as files and managed on data servers, and where infrastructures are built using IP network software. It is an industry where broadcasters need to control costs strictly and make the most effective use of their resources to enable them to remain competitive. The broadcast industry is ripe for a web/grid service based infrastructure because broadcasters are currently moving away from the conventional broadcast infrastructure and thus it enables a broadcaster to skip the conventional IT software infrastructure that

exists currently and make the most effective use of the inter-enterprise operation that is part of the broadcasting industry. The inter-operation with other broadcasters is increasingly an important part of a broadcaster's work flows because of the increasingly use of 3rd party production and production facilities.

The grid services in a media grid must be capable of meeting the very demanding reliability, reactivity, security and quality of service requirements that broadcasters require. A broadcaster operates under strict Government broadcasting requirements and an inability to meet advertised broadcasting commitments is a serious issue. An infrastructure must therefore include data replication and service replication to meet reliability requirements.

An infrastructure must be capable of reacting quickly to changes in a broadcast schedule. For example, if a programme overruns then the broadcast schedule and content required for that schedule must be moved quickly to the broadcaster. In addition, the necessary meta-data, accounting, auditing and legal authority must be managed if a broadcaster is to meet it broadcasting legal requirements.

The stored content that is to be managed by a media grid is large--25 Gigabytes per hour per channel currently which will increase to 100+ Gb/hour with high-definition programmes. The media industry has traditionally employed specialist high-speed file servers to enable the fast transport of data from server to broadcast location. (In the GridCast infrastructure the project has two BlueArc servers and each is capable of sustained 2 Gb/sec read and write speeds.) However, increasingly a broadcast infrastructure is a mixture of high-speed storage and conventional storage solutions. The grid services must therefore manage various levels of quality of service expectations within the broadcast infrastructure.

Each programme file is highly sensitive as it represents the Broadcaster's capital assets. The leaking of a story line of a programme, for example, may have a serious impact to the number of viewers for that programme. Within a broadcast infrastructure services must manage the conventional read, write and copy issues that are part of any data grid. In addition, a programme is subject to legal review and the maintenance of the legal authority with broadcast content is important issue in an infrastructure. Broadcast content will often be pre-processed prior to broadcast.

The GridCast media grid aims to provide a mechanism for sharing services and broadcast content and provide the basis for a prototype broadcast media service economy.

3. The GridCast Prototype Media Grid Infrastructure

The GridCast project has created a prototype media grid infrastructure, figure 1, to prototype media grid services. The infrastructure is intended to enable testing of public and private service provision, security policies, the reactivity of the services, the distribution of large-scale media archives and remote service provision. The infrastructure provides a private test-bed for the services that enable the management of broadcast schedules and the distribution of broadcast content to national broadcasters.

This media grid infrastructure uses the Janet network as a backbone infrastructure with private connections to Janet from BBC London, BBC Northern Ireland and between BBC Northern Ireland and the Belfast eScience Centre. The infrastructure has servers in each media site to host local services. In addition, specialist media processing services are available within BBC R&D and BBC Northern Ireland to enable prototyping remote service provision. Thus, for example, it is possible to manage broadcast schedules and distribute content to the point of broadcast within BBC Northern Ireland. The project uses the two BlueArc media servers to act as content repositories within the infrastructure.

To meet the needs of the BBC the network infrastructure is considered to be configured as defined in figure 2. An IP network provides a general-purpose communication infrastructure. A storage area network provides a means of content sharing that is designed to enable short-lifecycle content sharing such as that which is required for advertising trailers and news items. An ATM network provides a means of distributing high-quality live content with a definable quality of service. The grid service infrastructure must manage these distinct networks and enable their efficient management. Thus, for example if copying broadcast content between sites it will the most efficient mechanism required given the required quality of service.

4. A Broadcasting Scenario based Development

The GridCast project has taken a scenario based approach to its development and focused on developing grid services to implement the broadcast scenarios of interest to the BBC. These scenarios define a new way of running the BBC broadcasting network and consist of grid services inter-connecting the broadcast automation systems used in the BBC day-to-day operations.

The grid service infrastructure provides a way of linking together the automation and management systems that are currently used to in the development of broadcast content, the management of that content, the planning of broadcast schedules, the distribution of those broadcast schedules, the distribution of that content to support a schedule and the distribution of broadcast rights and meta data to enable broadcast schedule.

Thus, GridCast grid services must interact with

- broadcaster content and production management systems that monitor, catalogue and release programmes for broadcast;
- broadcaster schedule management systems that monitor, catalogue and release broadcasting schedules broadcasting network and the nations and regions;
- broadcaster playout (or transmission control) systems that manage content, that program local site automation systems and react to schedule changes in day-to-day operations;
- broadcaster transmission systems that prepare content for transmission;
- broadcaster archiving systems to record transmissions for future use; and
- broadcaster online distribution systems to service web based content access and delivery.

The broadcast scenarios have identified a broadcast content lifecycle and broadcasting workflows that must be managed and serviced.

4.1 A *Schedule* Scenario

A *Schedule* scenario links the distributed broadcast planning systems that manage and plan future broadcasts within the BBC network. A BBC network schedule consists of a backbone or Network schedule that is updated with a Nation's scheduled programmes such as local News and current affairs programme schedules. Grid services have been defined that enable the management of distributed schedules and that enable automated notification of schedule availability. Thus when a schedule planner in BBC Network finalises a schedule for a particular day the BBC Nations automatically receive local copies of that schedule. In general the media services provide a collection of schedule sharing services that allow the co-operative management of schedules for the broadcast locations.

4.2 A *Playout* Scenario

A *Playout* scenario is to broadcast the content for a particular television channel, i.e. control the internal automation systems that supply content to broadcast transmitters. The scenario begins with the finalisation of a schedule for a broadcast day in the schedule planning automation (as defined in 4.1). The finalisation of a schedule triggers a planning process to organise the transport of content to distributed broadcast sites. For most content this is a simple file copying process but for some content it will be subject to additional processing to create the required broadcasting format, to add subtitles or to add advertising trailers. All of these activities must be performed in a managed, secure, reliable and reactive service infrastructure. The scenario has many aspects that one would expect in a data grid but with additional security considerations to guarantee data integrity required by broadcast regulators.

5. A Media Grid of Services

The GridCast media grid provides a grid infrastructure that implements broadcast scenarios and workflows, and provides a dynamic multi-site GT3-based broadcasting grid. At the heart of this broadcasting grid are data management services that manage a distributed repository of broadcast content, provide a replica management system, provide automated content availability notification, automatic content copying, automatic content replication for resilience purposes, enable secure content copying between sites and implement content validation as required by broadcasting regulators. In addition, broadcasting schedule sharing services manage broadcasting schedule distribution, automated notification of availability, automated content copying planning service using a BBC defined network load model and automated content

copying. A collection of services is being developed to enable content processing as required by each broadcasting site.

The GridCast media grid services have been developed to enable the definition of sub-grids and virtual organisations within the GridCast media grid, see Figures 3, 4 and 5. (In GridCast the name Domain is used to denote a coherent collection of public services.)

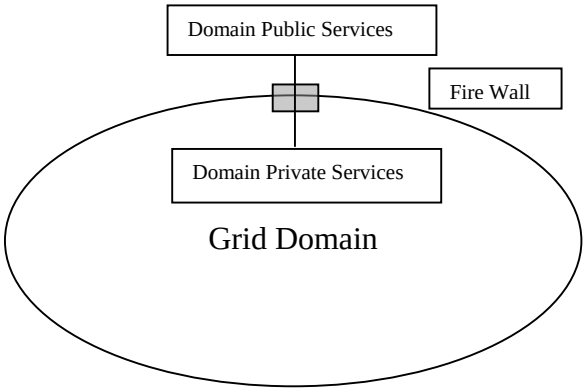

Figure 3: Domain based Service Structure

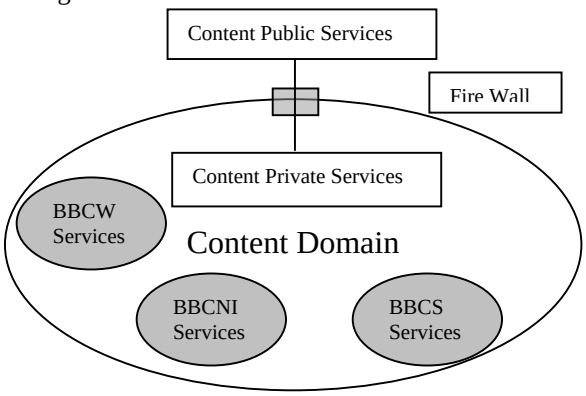

Figure 4: Content Management Domains

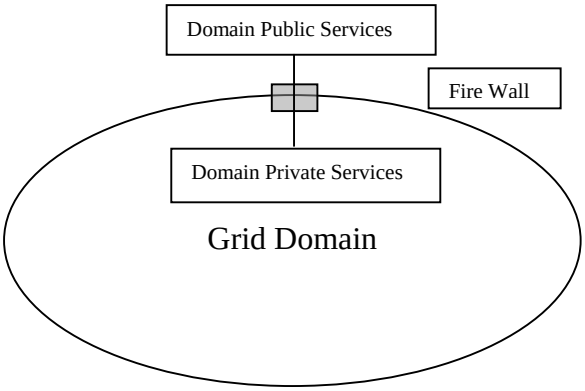

Figure 5: A Configuration Provider

A service is part of one or more domains. The software architecture enables multiple overlapping domains. The domains to which a service belongs are defined in the service configuration that a service retrieves from its configuration provider.

Thus, for example, all content management services might be defined to be part of a content management domain. A domain management domain would facilitate co-operation between the distinct content management services and give the illusion of a single content management service.

This architecture provides a flexible organisational service framework and would permit the dynamic formation of grid communities.

The configuration provider (CP) defines the structure of a domain. A configuration provider stores the configuration of the domain, such as the public service types, internal service brokers, the types of services in the domain, the configuration of all services, the valid machine names/addresses in the domain and the location of fixed services.

Any service in the domain must fetch its configuration from the CP and its type and location authenticated by the CP before a configuration is issued. All communications with the CP are encrypted and host and service certificates authenticated by the CP.

The CP provides a layer of security by enabling the scope of the domain to be defined and provides for the issue and exchange of domain certificates by the CP to authenticated services.

These services provide a broadcasting organisational infrastructure (such as BBC Northern Ireland or BBC Network), inter-organisational co-ordination (such as virtual organisations for content management or schedule management), distributed data management and remote service publication (that permits inter-organisation co-operation). These organisational structures permit definition of index services, broker services, discovery services, workflow services and security rules.

The GridCast grid services are intended to be lightweight and focus on authentication, discovery and remote service interactions. All long-lived processes are handled by custom broadcasting systems or workflow definitions

executed by a workflow processor. In defining our services as lightweight we mean that the service itself is short lived and intended to be fast; however, the grid container it sits within may be heavy weight such as that provided by GT3. This mechanism enables a variety of containers to be used from GT3 to simple soap interfaces.

In addition, each GridCast grid service has an associated web service which contains documentation on the service, a service log and (for some services) a collection of scenarios that can be performed from that service.

These web services are generated automatically using XSLT from the WSDL definitions for the grid services and built as part of the system deployment, see figures 7, 8, 9. The web pages are intended to provide a view of the grid services to enable evaluation of grid management issues in a large scale deployment.

6. Summary

The GridCast media grid has been deployed since November 2003 across three organisational networks.

The GridCast media grid represents a significant GT3 deployment with some 20+ service types and 30+ service instances in everyday use in the GridCast testbed.

The project thus far has concentrated on developing the necessary service infrastructure that is necessary to test the service deployment, performance and management issues. The next phase of the project will be to perform more intensive quality of service testing and service hardening. In particular, the project has yet to resolve the firewall issues that will be key to implementation of the domain structure.

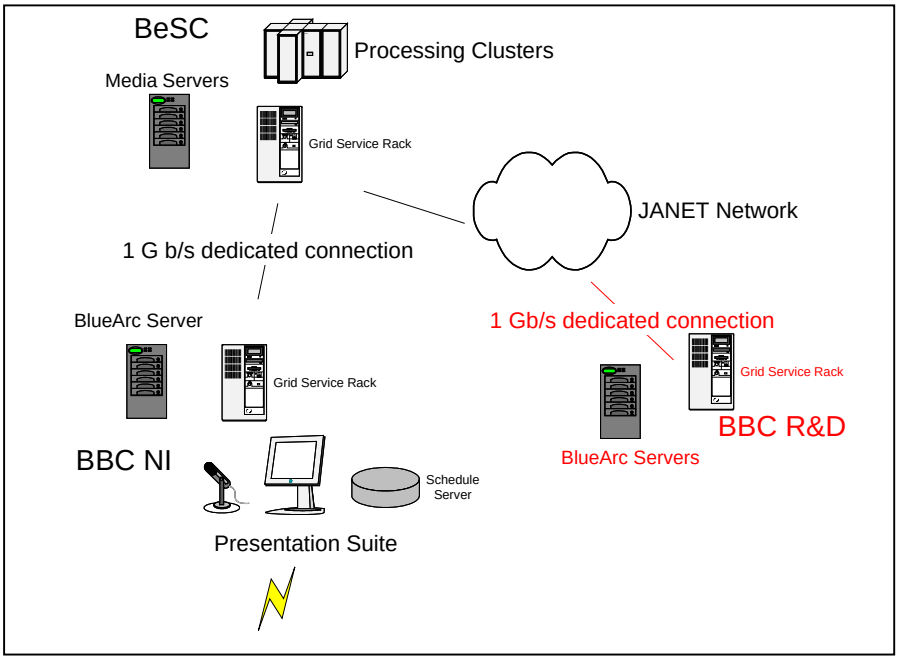

Figure 1: GridCast Infrastructure

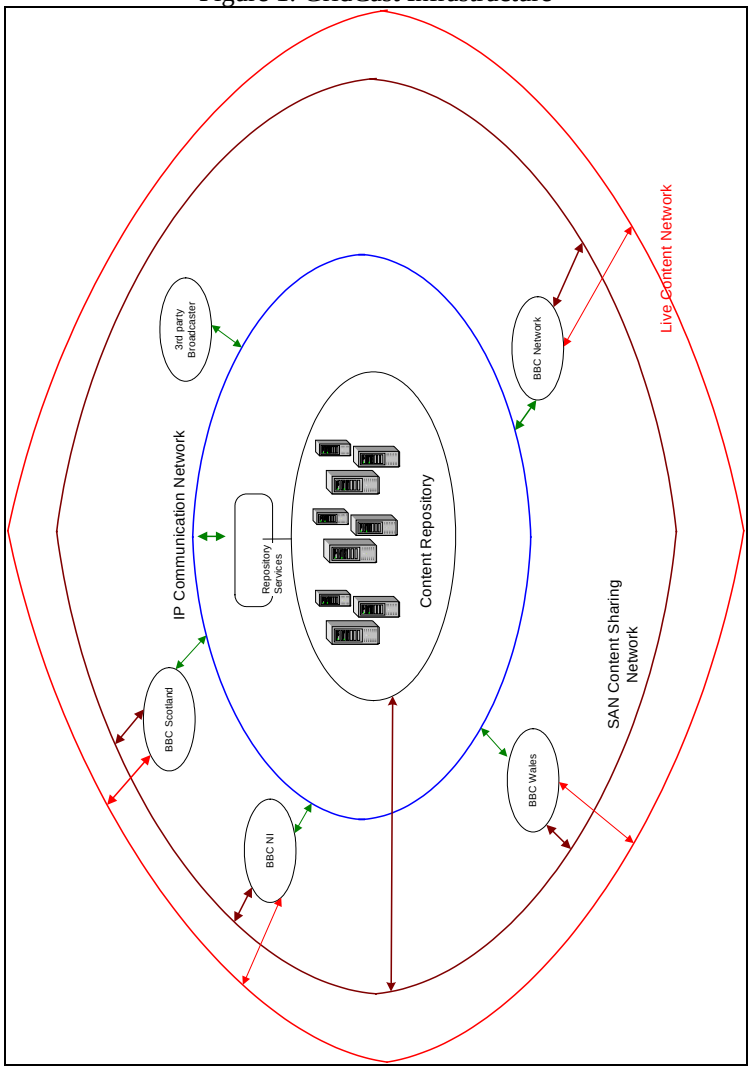

Figure 2: Network Topology

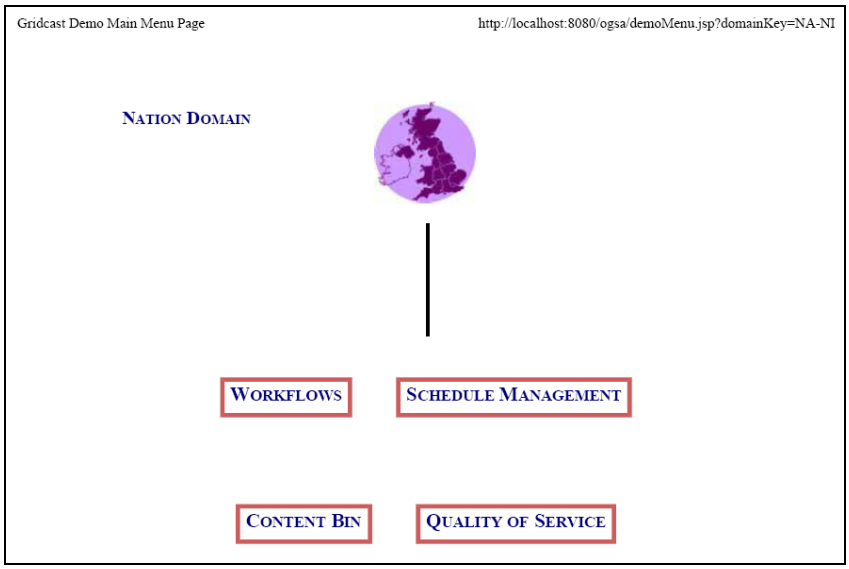

Figure 6: Nation Domain View

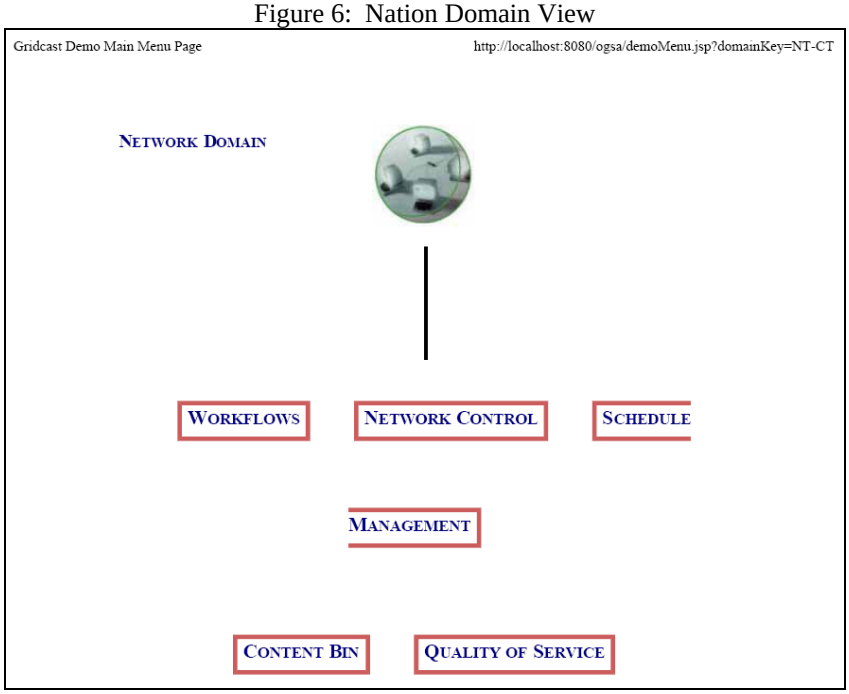

Figure 7: Network Domain View

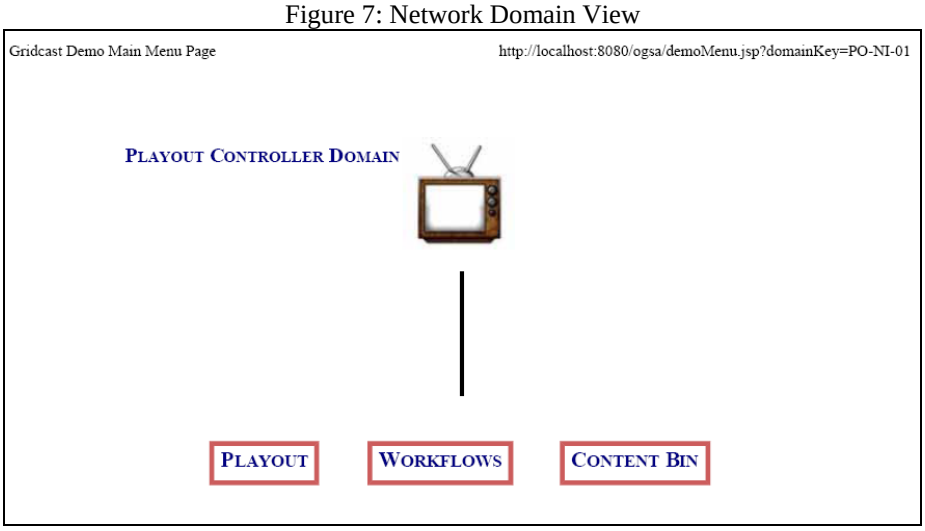

Figure 8: Playout Controller Domain