Using Metadata in SLA negotiation over Premium IP Networks

S. D'Antonio[†], M. D'Arienzo[‡], A. Pescapè[‡†], S. P. Romano[‡], G. Ventre[‡]

[‡]*University of Napoli "Federico II"*
Via Claudio, 21 – 80125 Napoli, ITALY

[†]*ITEM - Laboratorio Nazionale CINI per l'Informatica e la Telematica Multimediali*
Via Diocleziano, 328 – 80125 Napoli, ITALY

{maudarie, pescape, spromano, giorgio}@unina.it
{salvatore.dantonio}@napoli.consorzio-cini.it

Abstract

This paper describes a proposal to improve the negotiation of value-added services. The approach aims at supporting user-network interactions in an automatic fashion. Stemming from the concept of metadata, we designed a specific model for on-line delivery of digital multimedia resources with provisioning of guaranteed performance. We exploited such features in an architecture that provides network operators and service providers with the capability to perform a service creation and delivery process on top of a QoS-enabled infrastructure, with automated configuration of the underlying communication network. In this architecture neither the users nor the service operators need to have information about the QoS requirements of service content since the QoS related network operations are performed transparently and can be dynamically managed during the trading of the service.

1. Introduction

Service negotiation is most of the time seen as a user-centric process: the user needs to be aware of the requirements that a certain application or service has in terms of end-to-end communication performance. He is also involved in the service subscription process. While this approach finds the favor of users who desire a complete control over the services they are willing to subscribe, it requires also that users who have some skill in computer networks. On the other hand, many users simply desire to receive a service without any kind of intervention in its tuning. The solution that appears to be straightforward is that information related to content should be linked closely to the media itself, so that it could be made easily accessible to the users. There are some possible solutions to these issues that have appeared in the scientific literature. For example, in [1] some approaches are presented to allow a web server or a browser to ask for a certain QoS in the delivery of multimedia content. The first solution is quite naive, since it requires to insert, in the HTML document presenting a given resource, different links for each of the (predetermined) possible classes of service with which the resource can be downloaded. The second solution, instead, is based on the inclusion of QoS information within HTTP tags, so that a properly modified web server can use such information to negotiate with the network the communication service conditions. If also the browser were modified, then the QoS related information might be provided or modified interactively by the user. These two kinds of solutions do not link directly multimedia data and their corresponding QoS information since these data are simply stored in an HTML document in which a link exists to certain content. Therefore, we envisage problems with possible modifications that can be imposed on the QoS information since any change in it has to be immediately reflected in all the HTML documents referring to that specific multimedia content. Another solution is based on the request for service differentiation via explicit packet marking. This solution does not ask for changes at the application level, but rather in the operating system at both the client and the server sides; therefore it appears to be less flexible since it would require an explicit signaling mechanism between the application and the operating system for setting dynamically the appropriate QoS parameters. A different technique that recently appeared in the literature is based on the automatic detection of QoS requirements using a Layer 4+ analysis of the data exchanged between source and destination [2]. In this proposal, network nodes with the capability of interpreting the content of higher layer protocol headers and payloads, are expected to intercept data descriptive information to request QoS assurance for that media flow from the network infrastructure. Such an approach, albeit of great interest since it does not require any information structure or data to be linked to the media, relies on the development of techniques for the solution of two major drawbacks: i) how to successfully and reliably determine the QoS requirements from data such as protocol headers, file extensions, or other implicit

Proceedings of the 24th International Conference on Distributed Computing Systems Workshops (ICDCSW'04)
0-7695-2087-1/04 $20.00 © 2004 IEEE

information; ii) how to build network nodes that can perform efficiently such computational intensive tasks on the huge number of connections that are expected to characterize future multimedia services in the Internet. The solution we present in this paper relies on *Metadata*, and on the introduction of such a concept inside an existing architecture for the creation and negotiation of enhanced service. Metadata are data about data; they provide a means to describe resources in a structured fashion. Such descriptions may then be used by many diverse applications, such as search and retrieval tools for digital resources available on the Web. The specific goal is the incorporation of QoS aspects within a metadata model for multimedia resources. We start from a metadata model that has been defined in [3], where information for the identification of digital multimedia content for applications like Video Distribution or Distance Learning has been extended to include data related to the requirements that a network provider should satisfy in order to offer a guaranteed communication service. The model (defined as GEMSTONES) includes the parameters required for establishing data connections with service guarantees over IP networks, where QoS is offered on top of a *IntServ* or of a *Diffserv* infrastructure. The idea which motivated our work is that this approach can be exploited in the definition of an architecture for the dynamic, on-line access to communication services over Premium IP networks, i.e. networks where services are offered with proper performance guarantees as we will discuss in Section 2. The rest of the paper is organized as follow. Section 3 discusses the main issues related to the definition, negotiation and activation of Service Level Agreements and the concept of content-based SLAs. An actual example of the applicability of these new concepts is shown in section 4, where we address the issues related to the deployment and management of multimedia content distribution. Section 5 presents the QoS extensions we introduced in the metadata for multimedia content, while section 6 gives details about some implementation issues we faced. Finally, section 7 provides some concluding remarks, together with a discussion of our proposal.

2. Premium IP Networks

With the term Premium IP network we identify an IP-based communication infrastructure where users and applications can require the set-up of a data transport service with adequate performance guarantees in terms of parameters such as delay, jitter, and data losses. Even though in the rest of the paper we will refer to specific examples of these networks, in this section we will introduce a general model for the dynamic creation and provisioning of QoS-based communication services on top of IP networks [10]. We identify a variety of actors: users, service providers, brokers between users and providers and network providers. The idea behind the concept of Premium IP Networks, is that such entities can interact in order to define new services and to have the network automatically configured in order to guarantee that such services will be delivered with proper performance. The entities we have introduced are linked by means of a number of interrelations, for which the following assumptions hold:

a) users have a high level view of the service: they simply ignore (or are not interested in) the technicalities associated to it;

b) brokers are not necessarily aware of the semantic of a specified service;

c) service providers might have no knowledge about the specific delivery architecture managed by network providers;

d) network providers, in turn, are totally unaware of the high level signification of a particular service: from their viewpoint, a service is considered as the delivery of information according to a predefined set of rules.

To be consistent with these assumptions, we have introduced three major components (Figure 1) that we believe are needed to supervise the dynamic service creation and configuration process: Access Mediator (AM); Service Mediator (SM); Resource Mediator (RM). The Access Mediator manages user requests to the system. It adds value for the user, in terms of presenting a wider selection of services, ensuring the lowest cost, and offering a harmonized interface. The source of the services is a so-called "Service Directory" database, from which the Access Mediator performs transparent processing of the available information. For example, in the case of a Video Distribution service, it can select the cheapest offer if a movie is available from more than one service provider, and it can notify the user as soon as a new movie becomes available that matches the stored user's profile. Its main role thus consists in assisting and easing the service selection process.

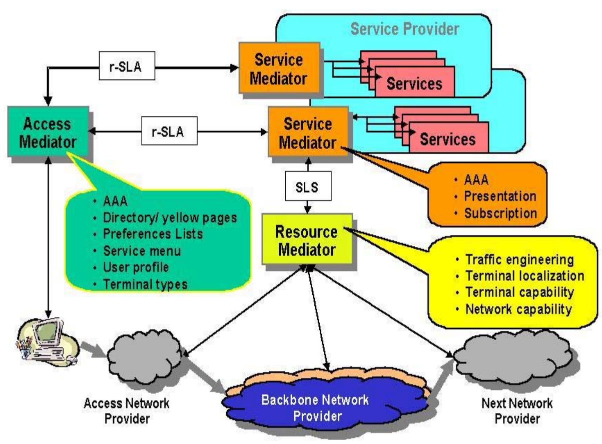

Figure 1: A Premium IP Network model

The Access Mediator may form associations with one or more Service Mediators, to which requests are issued. Generally off-line, the Service Mediator will take care of the creation of new services and of their presentation in the Service Directory. It is the task of the Service Mediator to map the Service Level Agreement from the Access Mediator into the associated Service Level Specification(s) to be instantiated in cooperation with the Resource Mediator(s). The SLA can be seen therefore as the interface between the Access Mediator and the Service Mediator. The interface between the Service Mediator and the Resource Mediator is the SLS, ensuring first of all the independence from both the high level view of a service and the specific network architecture employed. In a few words, an RM can offer all what can fit inside an SLS (and no more than this). Coming to the RM operation, SLS enforcement may be achieved by means of a policy-based approach, which ensures the correct operation of the network in a flexible and dynamic fashion.

3. SLA-based service negotiation

Keeping in mind the reference framework we presented in section 2, the scenario we envisage is one where users contact an Access Mediator (AM) in order to gain access to a number of value-added services, by means of negotiation of specific Service Level Agreements. The AM, in turn, needs to interact with one or more Service Mediators, each providing a certain set of services, to retrieve information about the characteristics of the services themselves. Afterwards, it organizes this information in order to let the user choose the service that most appropriately fits his needs. Once a specific service has been chosen, the involved Service Mediator(s) is (are) in charge of interacting with one or more Resource Mediators which, eventually, configure network elements so to efficiently satisfy the negotiated requests. This configuration might need to be performed considering the QoS-related characteristics of the service purchased. We therefore define content-based our SLA, since part of the parameters to be included in it are obtained directly from the multimedia content to be delivered. The process described above foresees the generation of a number of documents (Service Level Agreement, Service Level Specification, Policies), each describing the same instance of the service at a different level of abstraction and thus requiring creation/interpretation by the modules (Access Mediator, Service Mediator, Resource Mediator) belonging to the corresponding level of the overall architecture. Digging into the details of such mechanisms, we can see in Figure 2 that the Service Level Agreement is a contract between the end-user and the Service Mediator, negotiated via mediation of the Access Mediator. Once this contract has been proposed, the Service Mediator is in charge of translating it into one or more Service Level Specifications, containing a technical description of the service. This translation is a uni-directional process, requiring some additional information on the SM's side in order to retrieve, where necessary, service-specific data.

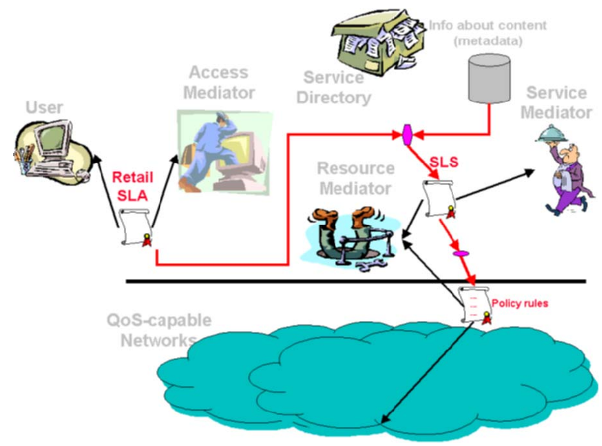

Figure 2: SLAs, SLSs, policy rules

The SLS is in turn given to the Resource Mediator, which, based solely on the information therein contained, has to become capable to perform admission control for the service bundle and, in case of success, to create the needed setup. The SLS represents a coherent and concise interface between the service-aware and the network-aware (though service-unaware) parts of the architecture. In a few words, it resides in a sort of limbo between the service and the network: it does not bear any more traces of the service it stemmed from, but at the same time it is still independent of the underlying network infrastructure. Finally, as already pointed out in section 2, SLS enforcement can be achieved by means of a policy-based framework: starting from an SLS instance, the RM has to derive the right policies to be installed on the network devices. In the next sections we will present how Service Level Agreements, and consequently Service Level Specifications can be properly formed by including QoS requirements of content or services by using metadata.

4. Linking Content to Quality of Service

Our metadata model is based upon a number of standard proposals in the field of on-line educational metadata, but we believe that it is general enough to be extended easily to other examples of multimedia content description. Attention has focused on the work done under the IEEE LTSC [4], the IMS project [5], CEN/ISSS LT [6], the ARIADNE project [7] and the Dublin Core [8]. Although not an educational initiative, the Dublin

Core aims at defining a set of cross-domain metadata to enhance the capability to search of electronic resources, and as such it has been the foundation onto which other domain-specific initiatives have been developed (The Dublin Core has recently formed the DC Education working group and the first working draft has been issued on February 2000.). The GESTALT courseware data model (GEMSTONES) is based upon IEEE's Learning Objects Metamodel (LOM) version 4.1, as adopted by both IMS and ARIADNE. GEMSTONES (GESTALT Extensions to Metadata STandards for ON-line Education Systems) organizes its metadata into eleven logically grouped categories [9], eight of which correspond to the baseline categories of the IEEE LOM model. The GEMSTONES additional categories are QoS, Assessment and Mapping.

The Quality of Service (QoS) category was introduced to map a learning resource's requirements to the capabilities of the network technology and/or services necessary for its delivery. This work is largely based on the IETF Integrated Services/ Differentiated Services models, but for the sake of compatibility with existing ongoing experiments and pilot infrastructures it has been designed to be adopted on top of ATM based infrastructures. Based on the previous considerations, a model for a possible QoS metadata structure is shown in Figure 3. It comprises both the ATM and the IETF models, while taking into account time as a new (yet optional) QoS parameter. As mentioned earlier, this structure is implemented in GEMSTONES.

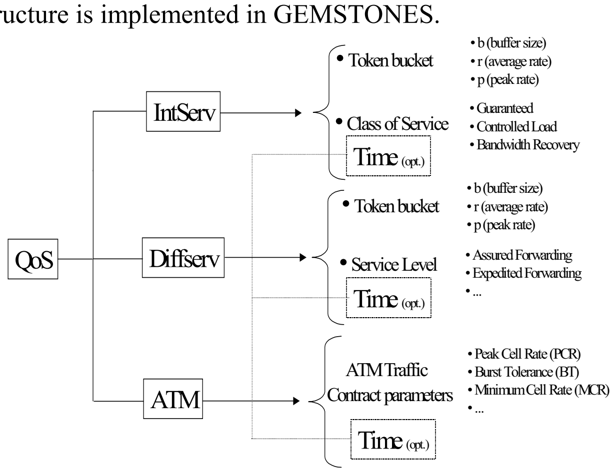

Figure 3: QoS metadata conceptual structure

5. An example of a Video Distribution service

In this section we will show how the generic architecture previously described has been implemented and how it can be exploited, via the definition of an appropriate business process, for the creation and delivery of a Video Distribution service, or, more in general, of Distributed Multimedia Applications. In our scenario, a user has access to the service via a user interface.

The GUI (Figure 4) for Video Distribution is implemented as a friendly web-based interface, which can be easily exploited even by users who are totally unaware of the technical details related to the service. As next section will disclose, while negotiating the service, the GUI is obtained as the outcome of a series of interactions among the various components of the architecture.

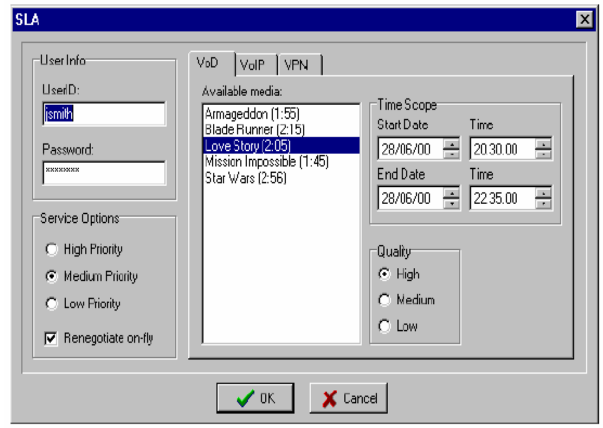

Figure 4: The GUI for a video distribution service

The template SLA is an XML file containing all the information necessary to uniquely identify the two parties, together with the service instance that has been negotiated; XML style sheets may be applied to this file in order to customize the way the SLA is presented to the final user. When negotiation starts, the user is required to indicate the service he is willing to perceive, specifying QoS level and, optionally, service lifetime. Again, we will summarize in the following the formal sequence of the steps needed at this stage:

– user subscribes to (or is authenticated from) the proper AM;
– user asks for negotiation of a new service instance;
– AM allows the user to choose one of the available services (in this case Video Distribution will be the user's choice);
– AM contacts a centralized repository in order to retrieve the service GUI associated with the selected service: in the simplified case we presented, such a GUI contains the list of the available movies, the time schedule for the service and the possible levels of QoS. Every movie title which appears in the GUI is available at least from one SM;
– service GUI is sent to the end-user.

At this stage, the AM does not make any semantic interpretation of the service under negotiation, but simply acts as a broker between the end-user and the Service Mediator. This has the advantage of relieving the AM from the responsibility of being aware of any specific service definition: the only entity involved in the definition process is, as one would expect, the SM.

Once received the service GUI, the user fills the required fields and submits his request to the AM. This event triggers the following actions on the AM's side:

- AM contacts all the SMs which registered as sellers of the specified service (in our case, Video Distribution). The list of such SMs may either have been obtained with the previous access to the repository (when the service GUI has been fetched), or be retrieved through a further access. A document containing the service parameters specified by the user is sent to the SMs in the list, in order to let them become aware of the service the AM (on behalf of the end user) is willing to receive;
- starting from the document just received, each SM creates one or more associated Service Level Specifications (SLSs) which are delivered to the right Resource Mediator;
- RM, based on such SLSs, makes an evaluation of the impact that the service is going to have on the network and translates it in the form of a 'cost' to be paid for service enforcement: such a cost is returned to the SM;
- SM is now capable to formulate an offer, which is sent back to the AM: the offer comprises a contribution coming from the cost information provided by the RM and an additional fee related to its own value-added service (e.g. content provisioning, brokerage activity with respect to network configuration, management of service options). In case of unavailability of the service, the quotation might contain an infinite cost value;
- once all of the quotations coming from the SMs have arrived, the AM sorts them according to the user's preferences, which may be derived from the user's profile. The sorted list of available offers is presented to the user: each single offer is built on the basis of the standard SLA template defined during the service creation phase;
- the user selects the offer which he deems most suitable; this operation, which has a legal value, is in all respects equivalent to the signature of a formal contract (SLA).

6. Implementation Issues

Some of the steps mentioned above require complex interactions among the various framework entities. In particular, SLA $\rightarrow$ SLS translation implies that, besides the information provided by the user, a number of different facets are taken into account, with special regard to the characterization of the traffic that is going to be generated upon media transmission and the features of the application used for streaming (joint vs separated video/audio channels, transport protocol utilized, type of encoding, buffer sizes). In the framework we defined, the translation from an SLA to the corresponding SLS(s) is quite straightforward, thanks to the modular decomposition of roles and responsibilities. Information available at the SM's side is interpreted in the light of the data carried inside the candidate SLA proposed by the AM: an interpretation is needed in order to build the appropriate Service Level Specifications. This derives from the consideration that an SLS is a technical document, differently than the SLA, which is an user-oriented service description: it is thus necessary to fill the gap between these two complementary perspectives on the concept of a service. Coming to the Video Distribution example, we will show in the following how the translation process takes place: as already stated, what we're going to present is specifically related to the SLS definition, whose main fields are:

- Scope, containing the network ingress and egress points for the traffic flow;
- Flow Identification, providing rules useful for the classification of the packets belonging to the traffic flow (source/dest network addresses, ports, transport protocol, etc.);
- Traffic Envelop, characterizing the traffic that the service is going to generate, in order to allow the network to both reserve the necessary resources and recognize the out-of-profile traffic;
- Excess Treatment, giving instructions about how to treat out-of-profile packets (marking, shaping, dropping);
- Performance Guarantees, containing the requirements imposed to the traffic flow (delay, jitter, throughput, packet loss);
- Service Schedule, specifying the time schedule over which the service has to be exploited;
- Reliability, defining the classical parameters related to service reliability (Mean Down Time, Time To Repair, etc.).

As a preliminary consideration, we notice that, depending on the application requirements, it might be necessary to translate the SLA in more than one SLS. This might happen, for example, when the application needs a duplex channel to work properly: one way to reserve resources (for the streamed multiplexed audio/video content) on the path from the video server to the end-user's system and the other to cope with streaming control data flowing in the opposite direction. This is true for the most common streaming protocols (RTP/RTCP [11], modified UDP versions [12], etc.) available nowadays. The situation obviously changes in those cases where the application needs to reserve completely independent audio and video channels (thus requiring one SLS each) or, stated in more general terms, whenever it is desirable to make a reservation for multiple, separate flows. It could be necessary to create more than one SLS also when different guarantees are to be assured over different time intervals. As to the traffic characterizations, they are expressed in the form of a sequence of time slots and related QoS parameters, generally in the form of a token bucket. Such characterizations are then represented inside our metadata schema: in this way information is linked closely to the media and becomes easily accessible to the Service Mediator in an automatic fashion. This approach of integrating metadata with multimedia content for guaranteed delivery of digital resources proves extremely successful: even for this step, users don't need to know anything about the communication requirements for the delivery of a certain multimedia document. All the work related to the negotiation of QoS guarantees with the network infrastructure can be managed and performed transparently by the mediation entities.

7. Discussion and Conclusions

This paper has showed how metadata describing multimedia contents or services can be used in a Premium IP network architecture to enable automated retrieval and QoS-guaranteed delivery of resources. Compared with other recent proposals in the area, our approach seems to offer several major advantages. Some solutions do not link directly multimedia data and the corresponding QoS information, since these data are simply stored in an HTML document in which a link exists to particular content. Second, possible modifications that in time can be imposed on the QoS information can cause problems in the automatic update of such data with the media itself, since any change in it to be effective has to be immediately reflected in all the HTML documents referring to that specific multimedia content. In our proposal, applications and browsers would make reference only to the metadata description of the QoS requirements that should be kept coherent with the related media content. Other solutions are linked to modifications to the operating systems on the web servers and on the client applications, since they generally require the development of explicit signaling mechanisms between the application on one side and the operating system on the other for setting dynamically the appropriate QoS parameters. Our solution does not rely on any change in the operating system of either client or server systems; additionally, it is more reliable and effective than those solutions recently proposed that are based on the automatic detection of QoS requirements from the interception of higher protocol header and payload information in ad-hoc developed network nodes. The trials we performed on a laboratory testbed demonstrated the feasibility of the approach of integrating metadata and user profiles technologies with multimedia applications, authoring tools and network protocols, for QoS guaranteed delivery of digital resources. The cost for this is that of characterizing the multimedia resource in terms of (a limited set of) parameters expressing the requirements on the network (and possibly on the client computer), to be provided to QoS-aware network equipment. It is felt that with this approach users don't need to know anything about the communication requirements for the delivery of a certain multimedia document. Of course, the more precise the media characterization, the more effective can be the usage of network resources. But, in our solution, this characterization should be therefore considered an authoring issue, rather than a end user responsibility.

8. Acknowledgements

This work has been carried out partially under the financial support of the "*Ministero dell'Istruzione, dell'Università e della Ricerca"* (MIUR) in the framework of the FIRB Project "Middleware for advanced services over large-scale, wired-wireless distributed systems (WEB-MINDS)".

9. References

[1] H. Ritter, T. Pastoors, K. Wehrle, "Diffserv in the Web: Different Approaches for Enabling Better Services in the World Wide Web", in Networking 2000, IFIP TC6 and European Commission International Conference, G. Pujolle, H. Perros, S. Fdida, U. Korner, I. Stavrakakis eds., Lecture Notes in Computer Science, Vol. 1815, Paris, France, May 2000.
[2] H. J. Stuttgen, Layer 4+ QoS Detection for access routers, AIMS 2000 Workshop, Heidelberg, October 2000
[3] The GESTALT Project, http://www.fdgroup.co.uk/gestalt/
[4] IEEE Learning Technologies Standards Committee (LTSC), http://grouper.ieee.org/groups/ltsc/
[5] The Instructional Management Systems (IMS) Project, http://www.imsproject.org/
[6] CEN ISSS Learning Technologies Workshop, http://www.cenorm.be/isss/Workshop/lt/Default.htm
[7] ARIADNE Educational Metadata Recommendation Summary, Version 2.0, April 1998
[8] S. Weibel et al., Dublin Core Metadata for Resource Discovery, IETF, Request for Comment, RFC 2413, http://www.ietf.org/rfc/rfc2413.txt, Sept. 1998.
[9] P. Foster et al, "The GESTALT project: Courseware Metadata Design (GEMSTONES)", Project Deliverable D401, http://www.fdgroup.co.uk/gestalt/D0401.pdf
[10] G. Cortese, R. Fiutem, P. Cremonese, S. D'Antonio, M. Esposito, S.P. Romano, A Dioconescu, "CADENUS: Creation and Deployment of End-User Services in Premium IP Networks", IEEE Communications Magazine, pages 54-69, Vol.41, No. 1, January 2003
[11] IETF Audio-Video Transport Working Group, RTP: A Transport Protocol for Real-Time Applications, IETF Proposed Standard RFC1889 (January 1996)

[12] University of Berkeley, Berkeley Continuous Media Toolkit — CMT, available at http://bmrc.berkeley.edu/frame/research/cmt/