Ladies and Gentlemen, Goodmorning! My name is Dimitris Oikonomidis and I am master of Science in Information Systems from Athens University of Economics and Business. First, I would like to give my thanks to Ericsson for giving us the opportunity to present our research work to a wide audience and of course for this wonderful trip to Sweden and its Development center. Now, I am going to make a short presentation of my 2003 Ericsson Award of Excellence winning thesis about economical aspects of Internet Content Distribution.

The outline of the presentation is as shown in the slide. First the problem of Internet Content Distribution is described and we make a review of the technical solutions that have been applied in order to deal with the incapabilities of Internet as a large-scale content distribution vehicle. Consecutively, we make a brief introduction to content distribution business models and next we focus on the economic dimension of the cache storage space allocation problem.

In the first years of the appearance of Internet and later the World Wide Web, their use was mainly scientific and educational. However, as millions of new users gain access to Internet and it is used as a medium for massive content distribution replacing, thus, traditional media like newspapers, television, new challenges are posed such as:

- Broadband Internet Access Technologies (Cable & DSL modems)
- Supply of rich in content dynamic pages and whole applications over Web
- Reliable delivery of streaming content.

This phenomenal growth of Internet disclosed the obstacles stemming from its client-server architecture and its best-effort nature.

Scarcity of resources, such as the available network bandwidth, that leads to High Transmission Costs. Overloading of content servers with Delays in creating and delivering dynamic content, congestion of network links, content unavailabity, response delay are some common symptoms that result to reduced end-user satisfaction.

One of the techniques that have been applied in order to alleviate these problems is caching technique. Storage nodes for temporary storing of popular content are deployed in the network *edge* near to end users. Caching takes advantage

of the *temporal locality* phenomenon, which characterizes content requests over the Internet. In this way, Content Requests to Origin Servers are reduced, the network core is relieved and packet traffic is diverted to network periphery so that download times end users experience are reduced dramatically.

Caching can be applied either in the form of **Full Replicas (mirrors), where we have** Static replication of one's provider content, or in the form of **Client – Side Caching, where** popular content from many providers is stored in shared caches in a Demand – Based Dynamic manner, that is content is replaced in caches accordingly to its popularity.

Content Distribution Networks CDNs are a cooperative scheme of caches operating under the supervision of a central administrative entity, the content distributor. CDNs act as a virtual overlay network of caches on top of the existing Internet infrastructure. The content distributor is responsible for the proactive replication of content in the storage nodes of its CDN, as well as for the intelligent redirection of end-users to the optimal node which contains the requested object. This redirection is based on several factors, such as geographical constraints, the links state and the network congestion.

The typical architecture of a CDN is depicted in this slide.

The chain of Internet Content Distribution is consisted from the following key players:

- Content Providers, who want their content to be easily accessed from end-users,
- Hosting Providers (Data Centers), who rent space to CDN to place their cache nodes
- Backbone Providers, who transmit the content from origin servers to cache servers or even end-users.
- Internet Service Providers (ISPs), who give end-users access to Internet and want popular content to be near to their customers,
- Content Distributors, who operate CDNs
- End – Users

Typically, the customer of a CDN may be a content provider who pays for his content to be stored near the end-users, and is charged by the content distributor for this service. The content distributor consecutively attributes a part of her revenue to third business entities which contribute to the content delivery service, such as the data centers and the backbone providers. We note that usually there is a peering agreement between ISPs and CDNs, so that CDNs place their caches free of charge in the premises of the ISPs, who are favored since their customers (end-users) can easily and cheaply access popular content from CDNs caches.

Let's focus now on the economic dimension of the cache storage space allocation problem. Assuming Content Providers as System Users, Cache storage space can be regarded as a scarce network resource. Demand for cache space, that is content requests to be served by a cache, exceeds its supply. The objective is to efficiently allocate cache space to users in such a way that social welfare is maximized. This can be achieved by caching Content with higher declared utility. Then Users (Content Providers) must declare their service valuation. But It is in users' interest to inflate their valuations in order to maximize their own utility. So, It comes in surface the Need of setting the appropriate Incentive Mechanisms so that it is in the users' interest to declare their true valuations of the caching service. In this way, it is assured that the most popular and "useful" content is stored in cache storage space in order to maximize the total utility that is extracted from the consumption of the service.

We develop an economic model about the market of storage space in web caches. Users (content providers) of the system compete in order to place their own content to the limited-size cache space. User i receives utility $\boldsymbol{U}$ when consuming x units of storage space. His utility function is the product of the hits (the content objects that are served by the cache instead of the origin server) multiply the Value per hit (given by content provider) is the sum of 2 dimension **:** his benefit from faster content delivery (increased end-user satisfaction, advertising income) and the bandwidth economies for not using network backbone for his content delivery.

Let's examine the economic mechanisms that have been proposed for the solution of the cache space allocation problem. Their incentive mechanisms are based either on an auction holding for storage space allocation or the provision of a differentiated caching service, where users self-select the class of service that

maximizes their utility, depending on the level of quality of service and the corresponding tariff.

In the case of the auction, we have an Auction holding per time spaces λ for storage space purchase.

- Content providers bid separately for each one of their content objects which they desire to cache
- Bids for content objects are of form { Bj **object value per byte**, Sj **object size in bytes** }. That is, The amount they finally pay for the caching of a j content object is the product of these two dimensions.
- In the end of the auction bids are sorted in descending order by object value per byte Bj.
- Clearing Price *p* is the highest losing bid (Case of Second Price Vickrey Auction).
- Content Providers are Charged in advance for proactive caching of their winning content objects and their guaranteed storage in cache until next auction holding.
- Result: Content Objects with higher marginal utility are placed in cache

In the case of a differentiated caching service, where users self-select the class of service that maximizes their utility, depending on the level of quality of service and its corresponding tariff. The motivation behind the Development of a differentiated caching service is the fact that Users do not valuate caching service in a homogeneous fashion. So we can have different classes of content (gold content, silver, bronze content) accordingly to their provider's valuation. Each class of content enjoys different Quality of Service based on the Hit-Rate, which is the percentage of successful responses from cache to the total number of requests. We can set an Objective of the system so that the Hit-Rate of the High class is three times the Hit-Rate of the low. We can achieve this objective by allocating different proportions of cache space to classes of content (e.g. 70% of space to high class, 30% of space to low class). Apparently, we also need a feedback mechanism to adapt deviations from objective by readjusting space allocation (e.g. 70% → 80% of space for high class)

We can Offer Incentive to users to select class that maximizes their utility by publishing a price *p* per *hit* different for each class of content.

Some final conclusions that can be made is that we must Insert user valuations of caching service in content delivery system in order to maximize social welfare.

Vickrey Auction Holding has the advantage that Offers incentives for true valuations but also the disadvantage that Inserts complexity in system (based on holding frequency)

Differentiated Caching Services Plus has the advantage that Satisfies needs of heterogeneous groups of users but also the disadvantage that the necessary Feedback Mechanism places an additional burden to the system

Ending, I feel the obligation to thank my supervisor professor Dr. C. Courcoubetis as well as the PhD candidate Mr. P. Antoniadis for their guidance, support and valuable advice that they gave me during the authoring of this thesis.