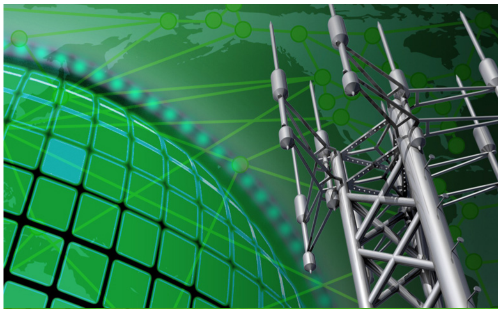
Storage. Networking. Accelerated.™

Whitepaper



# The Business Case for Caching in 4G LTE Networks

Prepared for LSI

By Haig Sarkissian - Wireless 20|20

Wireless 20|20

## Table of Contents



In order to deliver exceptional user experience, LTE networks must reduce **response time** to a fraction of a second.

## 1. Executive Summary

Wireless network operators are adding bandwidth with 3G and more recently 4G LTE technologies to meet the growing data demands of smartphone users. This bandwidth will provide coverage and capacity; however, it does not address the need for low latency.

Data providers like Google recognize that low latency is a key requirement for their users. They believe that web page response times need to be in the range of 0.25 second in order to deliver exceptional user experience.[1] Today's LTE data networks are delivering 7-20 seconds of response time.

Wireline networks have used content delivery networks (CDNs) and caching extensively in the Internet in order to reduce response time. This paper analyzes embedding cache in wireless networks to achieve significant reduction in response time and thus provide exceptional wireless user experience.

The business case for CDNs in large wired networks has been well established. This paper further analyzes the business case for caching in wireless networks. Because caching reduces backhaul traffic, particularly peak busy hour traffic, the financial benefits of implementing caching in the core network as well as at the edge of the network were investigated. Using the WiROI™ Tool, an industry-recognized LTE business model, we calculated and translated the peak busy hour traffic reduction into OpEx savings for three sample markets.

The results show that implementing Level 2 caching in the core network will result in savings of the Internet connection fees. Implementing Level 1 caching at the edge of the network offers significant backhaul traffic reduction that translates to sizeable cost savings in both OpEx and TCO. The analysis showed that network operating cost savings ranged from 29% to 35% and TCO savings ranged from 8% to 11% over the ten-year study period. Since backhaul transmission expense was the highest cost element in the TCO for all three markets analyzed, the largest amount of savings came from the traffic reduction due to the Level 1 cache in the base station.

## 2. Benefits of Caching LTE Traffic

Caching is a mechanism for temporary storage of mobile data traffic, including HTML pages, images and video in order to reduce backhaul bandwidth and server load. The main benefits of caching are to reduce response time and improve network efficiencies.

### a. Faster Response Time

Response time is used to describe the time it takes for a mobile broadband user to request information by clicking on their device until getting online, opening a web page or receiving the first frame of a video clip. Hence response time includes the time to encode the packet for transmission and transmit it,

*(cont. p.2)*

[1] *http://www.nytimes.com/2012/03/01/technology/impatient-web-users-flee-slow-loading-sites.html?pagewanted=all*

plus the time for that data to traverse the access network equipment, including the base station, backhaul, transport and the core network. Once at the core network, the request often needs to be served over the Internet by a server at a remote datacenter.

Caching improves OpEX by reducing total throughput, and improves CapEX by reducing peak bandwidth required.

This Internet latency is highly variable and depends on the proximity of the data center to the mobile network operator's point of connection to the Internet. Once at the datacenter, the request is subject to further processing by the server, including data gathering and checking. Although a minimum bound on latency is determined by the distance between communicating devices and the speed at which the signal propagates in the circuits, actual latency is often much higher because of packet processing by the Internet networking equipment. Data sent by the server is often subject to further data filtering, virus checking and scrubbing for unwanted information prior to deliver to the mobile device.

The goal of caching is to bring popular content closer to the mobile user by intelligently duplicating it in the core as well as at the edge of the LTE network. This process practically eliminates the variable latency of the Internet and significantly reduces server response time by assuring that the content of the cache is pre-scrubbed and ready for delivery.

**b. Reduced Backhaul Traffic**

Advanced caching techniques promise to reduce the overall backhaul traffic and improve network efficiencies. By incorporating large

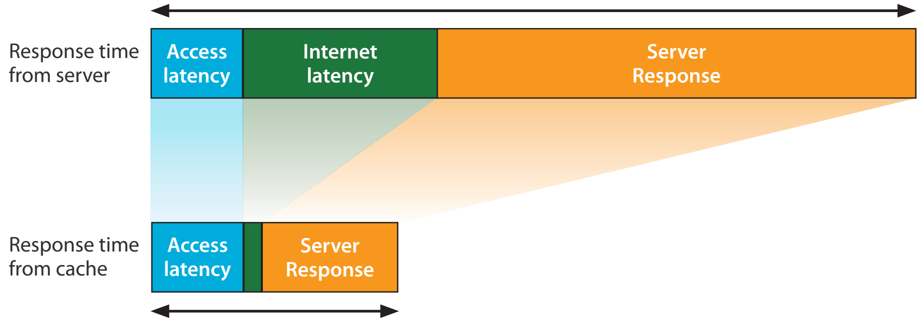



Figure 1: Improvement of Internet Latency with Caching

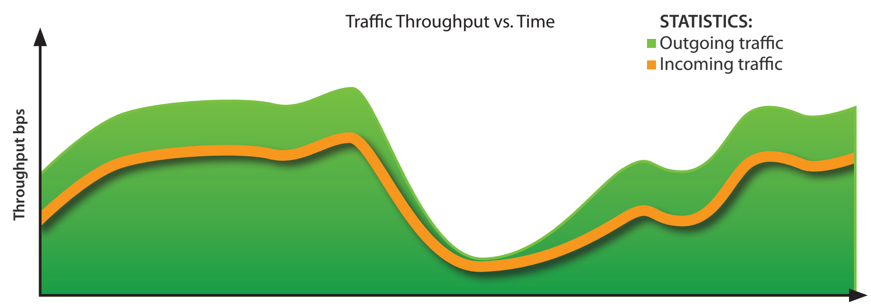



Figure 2: Improvement in Peak and Average Traffic with Caching

amounts of storage (64GB-256GB) in the base station, user data can be pre-fetched and stored at the edge of the network for instant access by the base station scheduler.

This data is often gathered during idle timeslots on the backhaul network, thereby significantly reducing busy hour traffic.

According to industry sources, up to 40% of backhaul traffic can be reduced at the busy hour.[2] This significant reduction in busy hour traffic can translate to measurable cost savings of recurring monthly backhaul operating expenses. News events such as sporting bring many users to similar content in a short period of time, thereby creating traffic spikes. Caching at the edge of the network has an effect of smoothening of traffic spikes and averaging the backhaul traffic over a longer period of time.

**3. Caching in the LTE Network**

Caching can be used by mobile network operators to store and serve popular content

Caching close to the end user improves user latency by reducing server response time, and almost eliminating Internet latency.

in the network in order to provide instant access of this content to mobile subscribers. Initially, caching is being implemented in the core network as operators take a centralized approach to caching. This reduces the uncertainties of Internet latency and remote server response times.

As data traffic increases on LTE networks and backhaul links become congested, distributed approaches to CDN deployment are becoming practical, thanks to the all-IP architecture of LTE eNodeBs. These new techniques are promising further improvement in latency and network efficiency by pushing the cache to the edge of the network.

*(cont. p.3)*

*[2] From various sources. See, for example," should be inserted in front of the URL http://www.altobridge.com/2012/12/03/altobridge-proves-40-data-backhaul-savings-with-intelligent-content-delivery-in-commercial-wireless-networks/3*

**a. Caching in the LTE Core Network (Layer 2 Caching)**

Caching servers can be deployed anywhere within an all-IP network. In LTE networks, Layer 2 (L2) caching is typically deployed at the packet data network gateway or the PGW, within the evolved packet core (EPC). These cache servers offer multi-terabytes of data storage capacity and capture popular content that is accessed by a large number of users. The beauty and simplicity of deploying a centralized cache in the EPC is making them popular in LTE networks. They provide a mobile network operator with many of the benefits of caching. Centralized caching reduces the uncertainties of Internet latency and remote server response times. Because they have visibility of data seen by large number of subscribers, L2 caches will enjoy high hit ratios.

When a popular content is found in an L2 cache in the EPC, this information is no longer obtained from the remote server of origin. This reduces the traffic between the EPC and the Internet, thus reducing the bandwidth and cost of the Internet connection. Therefore, the financial benefit of L2 caching is limited to the cost reduction of the Internet transit.

**b. Caching at the Edge of the Network (Layer 1 Caching)**

The popular content in the example above still needs to traverse the transport and backhaul links of an LTE network. Often times, the most expensive part of the network operating cost is the cost of the various fiber leased lines that connect the LTE base station to its core network. Layer 1 (L1) caching brings this functionality as close to the end use as possible by placing it in the base station. LTE macro and metro base stations, and pico base stations are ideal platforms to incorporate L1 caching because of their all-IP architecture and their proximity to the end user.

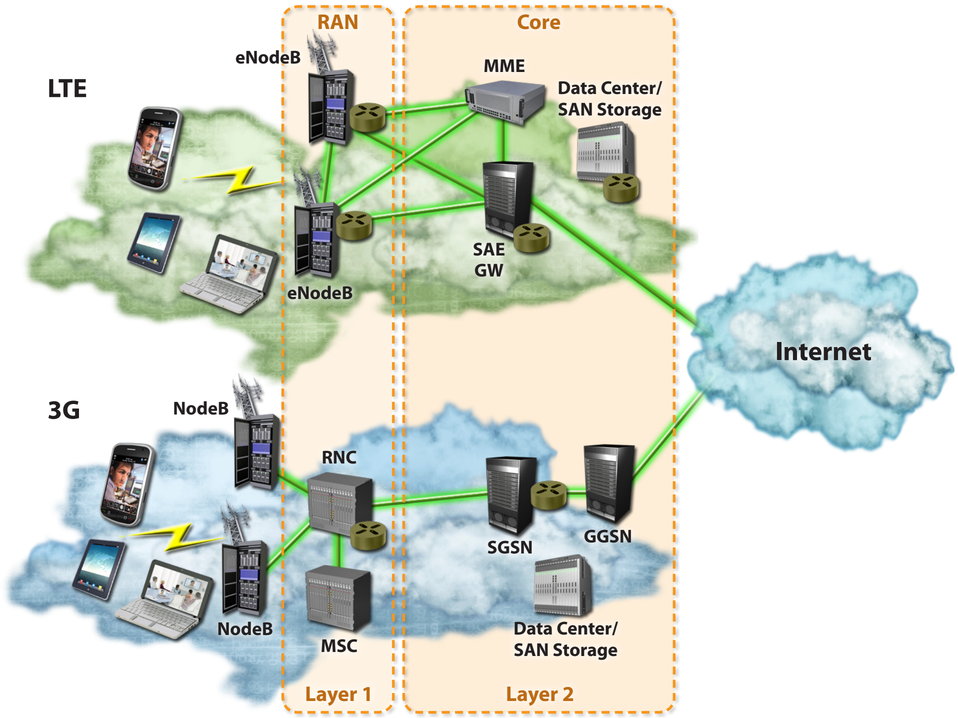



Figure 3: Layer 1 and Layer 2 Caching Locations in a Wireless Network

Caching at the network edge further improves caching in the core network.

Sophisticated caching algorithms, including proactive, predictive and coherent caching, coordinate information flow between the L1 and L2 caches in order to increase the probability of hits. Special video optimization algorithms further improve the buffering of video streams and coherently synch them with the base station scheduler in order to make optimum utilization of precious radio access network resources.

By placing caching in the radio access network (i.e., in the base stations), backhaul traffic is considerably reduced, and since backhaul and transport costs are often the largest parts of the LTE network OpEx, the financial savings can be significant.

## 4. Case Study Methodology

To study the impact of caching on an LTE service provider business case, Wireless 20/20 selected three network markets for the case study. These were Chicago, USA; Munich, Germany; and Hangzhou, China. Wireless 20/20 has developed over 60 business cases using the WiROI™ Business Case Tool for clients all over the world including these markets, and the network models, cost data and population metrics for each of these markets was generated using real-world data from these studies. Each of these cities presented a variation in population density, backhaul costs and mobile data usage, making them appealing and exemplary scenarios to analyze. A complete LTE network model was built for each of the three markets using the WiROI Business Case Tool. The output of the WiROI Tool, an investment-grade, 10-year business case including full income statement, was captured to establish the baseline CapEx, OpEx, TCO, NPV and IRR.

An L2 caching model was built to simulate the impact of L2 caching on the business case. CapEx for this system was incurred up front as well as during the course of the ten-year study period in order to accommodate the incremental storage and intelligence required as traffic and number of subscribers increased over time. The percentage of traffic savings from the Core Network connection to

(cont. p.4)

the Internet was used to calculate the Internet connection cost savings and to capture the associated network efficiency improvements.

Similarly, an L1 caching model was built where caching functionality was added in each eNodeB during the initial deployment of the LTE network. As the network was expanded to handle the increased capacity needs of the users, new eNodeBs were added equipped with caching. Both macro and pico base stations were used for coverage and capacity, respectively. The incremental cost of adding caching functionality to the base stations was calculated as part of the CapEx. Any traffic savings due to the L1 cache was captured in order to calculate reduction in backhaul and transport OpEx. These savings were calculated on an annual basis for the ten-year study period.

The WiROI Tool's easy-to-use graphical user interface dashboard allows the user to separately turn L1 and L2 caching on and off

> The cache ROI model allows a user to select the estimated costs and bandwidth reductions from caching.

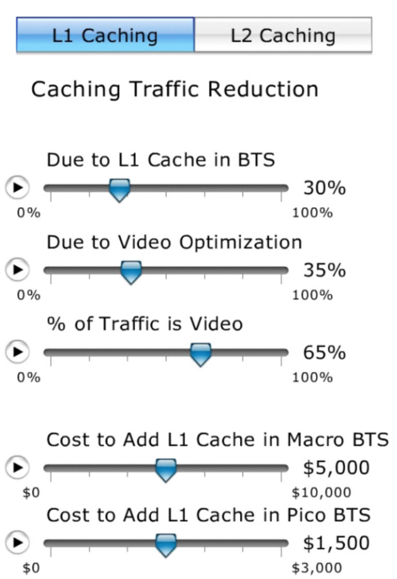



Figure 4: Cache Cost Sliders

and view the impact of each caching method on the LTE network. The two scenarios that were analyzed were an LTE network with L2 caching only and an LTE network with both L1 and L2 caching. The results were captured and then compared with the baseline case where no caching was used.

**a. CapEx and Technical Assumptions**

The cost to add L2 caching functionality to the core network was set to US$50,000, which provided caching capability to handle up to 1Gbps traffic at the busy hour.[3] The amount of traffic reduction due to the L2 cache was set to 30%. In order to allow for sensitivity analysis, sliders were set with a range on each of these parameters as shown in Figure 4.

The cost to add L1 caching functionality in a base station was set differently for a macro vs. a pico base station. For an omni-directional pico BTS, the price was set at $1,500. In the case of a three-sector macro BTS, the price was set at $5,000 since more memory and intelligence is needed to manage a larger number of users on a macro base station.

The amount of traffic reduction due to an L1 cache was set at 35% for video traffic and 30% for non-video traffic. In addition, 65% of the total traffic volume was assumed to be video. These assumptions are consistent with industry sources.

**b. OpEx Assumptions**

Since one of the main financial benefits of caching is due to the reduction of expensive backhaul and international Internet connection (IIC) costs, it was important to obtain accurate cost information for these OpEx parameters. These costs are typically paid on a recurring basis and are usually expressed in monthly terms for each leased line channel capacity. For example, a base station backhaul connection to a fiber DS3 line provides 45Mbps peak capacity and could be priced at $2,500 per month. Often, the mobile network operators lease a portion of their backhaul network from fiber network operators who are sometimes their competitors. These costs vary greatly from country to country and from operator to operator and are sometimes subject to government-mandated tariffs that change slowly. For example, the cost to connect to the Internet in an island country or an African

> The caching ROI model uses realistic costs for backhaul transport, but allows the user to change the costs for different scenarios or markets.

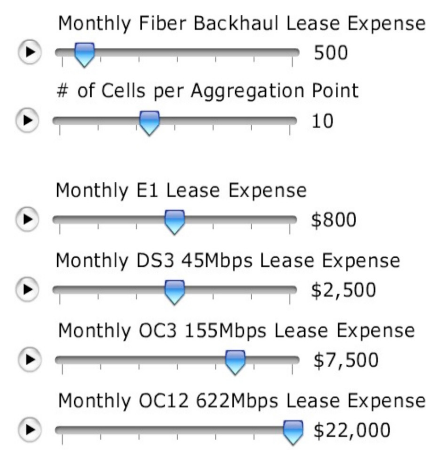



Figure 5: Backhaul Cost Sliders

country may be significantly higher than in Europe or North America. As traffic consumption increases, the cost of the backhaul and Internet connection is projected to increase.

A backhaul network for a large mobile operator can be complicated and usually not heterogeneous. An operator uses a mix of fiber, copper, P2P microwave and P2MP microwave connections as the first level of backhaul from their base stations. These links get aggregated at various aggregation points and the combined traffic is then carried by a fiber transport network to the EPC core where it is connected to the Internet.

Our model assumed a mix of microwave and leased line connections as the first-level backhaul from base stations, with 60% microwave and 40% leased line. This mix was presented as a slider on the dashboard, which allowed the user to vary this mix from 0% to 100%.

*(cont. p.5)*

[3] *Costs based on early estimates of cache equipment costs, but may vary from market to market. The WiROI model can be used to adjust such cost estimates for sensitivity analysis.*

**The caching ROI model provides detailed breakdowns for all aspects of network OpEx and CapEx.**

We also assumed that on the average every 10 base stations were connected with an aggregator. This parameter was also presented as a slider. Finally a scalable set of fiber transport channel options was set up to carry the aggregate traffic to the core network. These included DS3-45Mbps, OC3-155Mbps and OC12-622Mbps lines, each with its own monthly lease expense.

A least cost algorithm was set up in the WiROI Tool to calculate the lowest-cost method of transporting the traffic based on the busy hour traffic capacity needed. This was adjusted year over year to accommodate the increase in the busy hour traffic. Furthermore, an annual cost-reduction parameter was included to allow for such reduction.

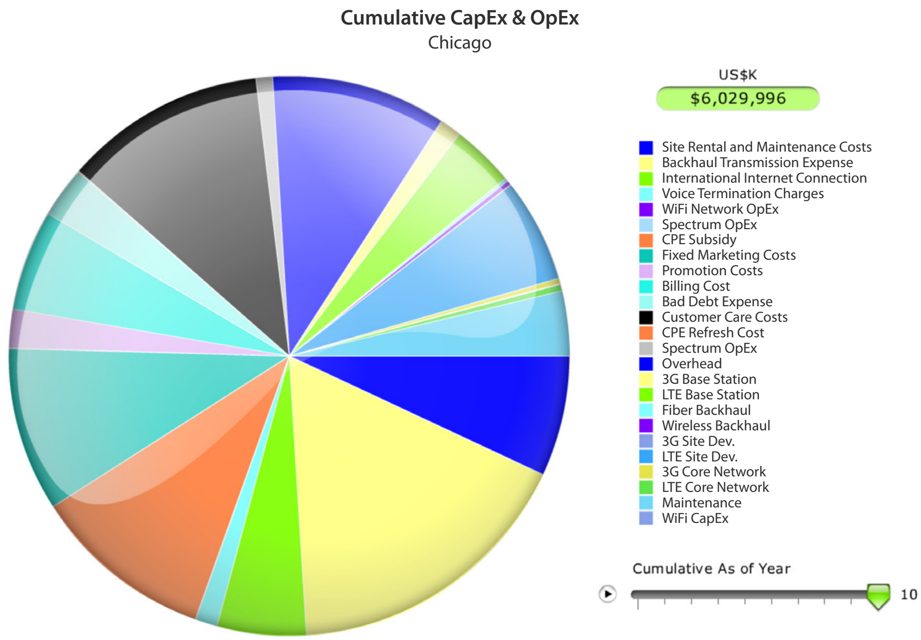



Figure 6: Detailed Cost Categories in the WiROI Model

## 5. Business Case Study Findings

The WiROI Tool calculates the total cost of a network and provides several financial metrics to evaluate the impact of caching on the network. Financial metrics include yearly and cumulative OpEx, CapEx and TCO, as well as IRR and ROI metrics. An example output from the WiROI model, showing ten-year TCO, is shown in Figure 6. Cost categories include all

**L1 and L2 caching combined shows the greatest improvement in TCO.**

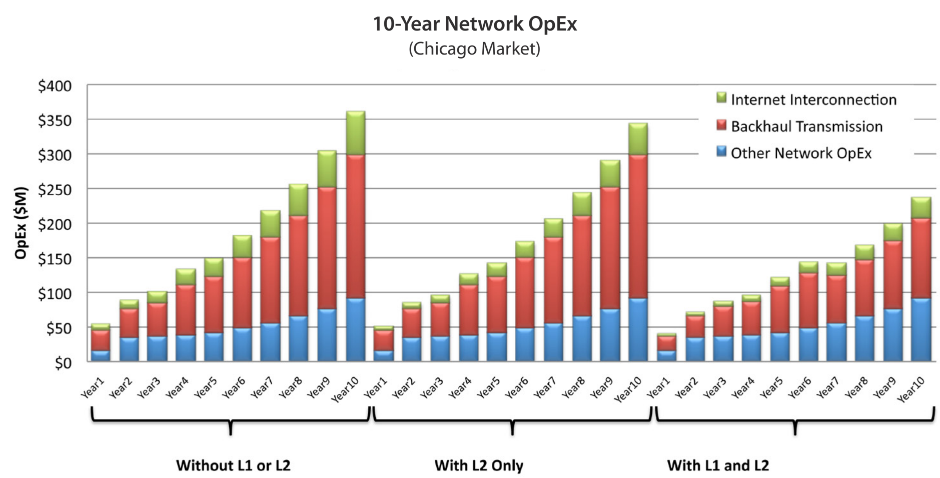



Figure 7: 10-Year Network OpEx

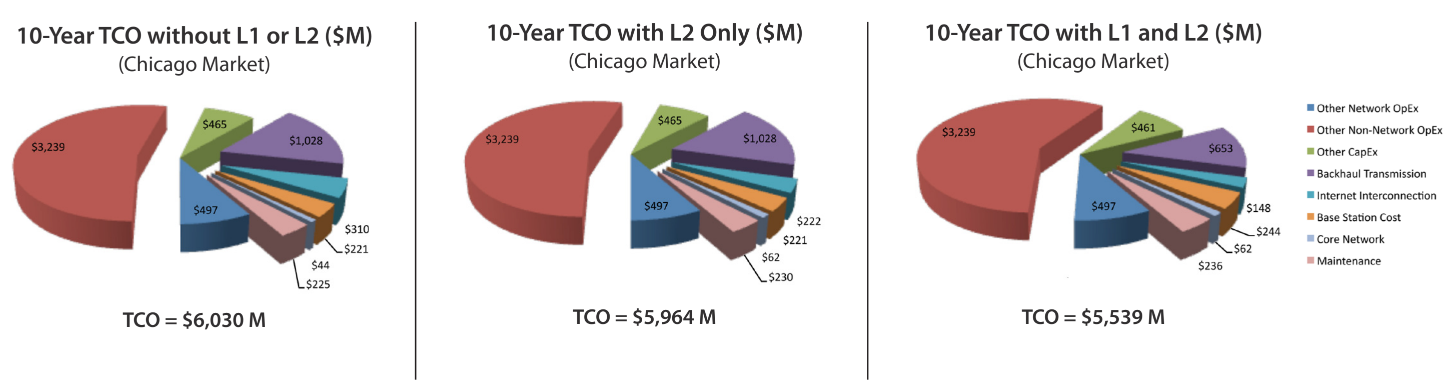



Figure 8: 10-Year Total Cost of Ownership

(cont. p.6)

aspects of building, maintaining, marketing and administering a network.

To study the impact of caching on the business case of an LTE service provider, the WiROI model was run for all three markets (Chicago, Munich and Hangzhou), with and without caching implemented in the network. Two types of caching were modeled: L2 caching alone (in the core network) and a combination of L1 caching at the network edge (eNodeB) with L2 caching in the core.

The results showed that network operating cost savings for combined L1 and L2 caching ranged from 29% to 35% and TCO savings ranged from 8% to 11% over the ten-year study period. The largest cost improvement came from reduction in backhaul transmission expenses, followed by Internet interconnection expenses, as shown in the Figure 7 (from the Chicago market).

While the cost of caching equipment increased the total CapEx, the benefit of lower network operating costs was far greater, resulting in a net decrease in total cost of ownership (TCO) over a ten-year period. The ten-year TCO for the Chicago market is shown in Figure 8. Backhaul transmission expense was the highest cost element in the TCO for this market, and the largest amount of savings came from traffic reduction due to the L1 cache in the base station. While some additional CapEx was required, due to the additional cost of caching equipment, the improvement in OpEx was far greater than the increased CapEx.

**a. Network OpEx Savings**

As shown in Tables 1 and 2, the case studies for all three markets showed the same pattern of significant savings in Network OpEx.

Implementing L1 caching at the edge of the network offered significant OpEx savings - ranging from 27% to 36% of total Network OpEx.

Table 1: 10-Year Network OpEx

| Chicago Market | No Cache | L2 Cache Only | L1 and L2 Cache | Improvement |
|---|---|---|---|---|
| Internet Interconnection ($M) | $310 | $222 | $148 | 52% |
| Backhaul ($M) | $1,028 | $1,028 | $653 | 36% |
| **Total Network OpEx ($M)** | **$1,836** | **$1,748** | **$1,298** | **29%** |

| Munich Market | No Cache | L2 Cache Only | L1 and L2 Cache | Improvement |
|---|---|---|---|---|
| Internet Interconnection ($M) | $164 | $117 | $78 | 52% |
| Backhaul ($M) | $276 | $276 | $172 | 38% |
| **Total Network OpEx ($M)** | **$569** | **$522** | **$380** | **33%** |

| Hangzhou Market | No Cache | L2 Cache Only | L1 and L2 Cache | Improvement |
|---|---|---|---|---|
| Internet Interconnection ($M) | $190 | $136 | $91 | 52% |
| Backhaul ($M) | $135 | $135 | $92 | 32% |
| **Total Network OpEx ($M)** | **$405** | **$351** | **$262** | **35%** |

All three case studies showed that implementing L1 caching at the edge of the network offered significant backhaul traffic reduction, which translated to sizeable cost savings in both backhaul and transport OpEx. Munich and Hangzhou showed slightly better results due to variations in the cost of backhaul assumed and the amount of traffic over the ten-year period.

Table 2: 10-Year Total Cost of Ownership

| Chicago Market | No Cache | L2 Cache Only | L1 and L2 Cache | Delta |
|---|---|---|---|---|
| Total CapEx ($M) | $955 | $977 | $1,002 | + 5% |
| Total OpEx ($M) | $5,075 | $4,987 | $4,537 | - 11% |
| **TCO ($M)** | **$6,030** | **$5,964** | **$5,539** | **- 8%** |

| Munich Market | No Cache | L2 Cache Only | L1 and L2 Cache | Delta |
|---|---|---|---|---|
| Total CapEx ($M) | $248 | $254 | $260 | + 5% |
| Total OpEx ($M) | $1,428 | $1,382 | $1,240 | - 13% |
| **TCO ($M)** | **$1,676** | **$1,636** | **$1,500** | **- 11%** |

| Hangzhou Market | No Cache | L2 Cache Only | L1 and L2 Cache | Delta |
|---|---|---|---|---|
| Total CapEx ($M) | $163 | $168 | $172 | + 6% |
| Total OpEx ($M) | $1,047 | $993 | $904 | - 14% |
| **TCO ($M)** | **$1,210** | **$1,161** | **$1,076** | **- 11%** |

The analysis further showed that implementing an L2 cache resulted in savings in the IIC charges that are typically higher for developing countries or countries far away from North America, Europe and developed Asia. For the three business case scenarios analyzed, the savings in terms of dollars due to L2 were marginal because the cost for the Internet connection was low. But even for developed countries, the percentage savings on the Internet connection cost was impressive.

*(cont. p.7)*

Table 3: Comparison of WiROI Market Models

| | Chicago | Munich | Hangzhou |
|---|---|---|---|
| Market Population | 9.5 million | 2.58 million | 1.8 million |
| Area Coverage | 7,688 sq. km | 1,510 sq. km | 296 sq. km |
| Initial Number of Base Stations | 1,938 | 603 | 113 |
| Number of Base Stations Added over Ten Years | 3,661 | 740 | 968 |
| Total Busy Hour Traffic – Year One | 30Gbps | 10Gbps | 8Gbps |
| Total Busy Hour Traffic – Year Ten | 460Gbps | 125Gbps | 100Gbps |

### b. Total Cost of Ownership

As shown in Table 2, the case studies for all three markets showed lower TCO over a ten-year period with caching. Adding L2 caching alone to the network showed an improvement, but the largest gain (both from absolutely numbers and percentage improvement) came from adding L1 caching at the eNodeB.

Chicago, being the largest market (population and area coverage), required the highest amount of initial and ongoing CapEx. Total cost to add caching to the Chicago market was likewise higher. However, for all three markets, the improvement of TCO was in the same range of 8% to 11%. Table 3 shows a comparison of the key parameters that affected the size and growth of the three study markets.

## 6. Conclusion

Transparent L2 caching, which allows operators to cache (or copy) traffic that frequently transverses their networks and store it closer to end users, has several benefits. The use of transparent caching means that popular content does not have to be repeatedly requested from central servers. This mitigates the need for network operators to purchase new IIC capacity, especially when end-user traffic requirements continue to grow.

So far, the obvious motivation for operators to deploy L2 caching has been cost saving on Internet connections, particularly international connections. But as new, sophisticated and yet cost-effective technologies enable the evolution of caching by moving it closer to the end user, L1 caching in the LTE base station offers even further cost savings on the backhaul network, which represents by far the much greater component of network operating cost.

In addition, the deployment of complementary and interoperable L1 and L2 caching offers LTE operators the opportunity to significantly improve the quality of experience they offer end users and open new doors for location-based, value-added services.

Storage. Networking. Accelerated.™

For more information and sales office locations, please visit the LSI web sites at: lsi.com lsi.com/contacts

**North American Headquarters**
San Jose, CA
T: +1.800.372.2447 (within U.S.)
T: +1.408.954.3108 (outside U.S.)

**LSI Europe Ltd.**
**European Headquarters**
United Kingdom
T: [+44] 1344.413200

**LSI KK Headquarters**
Tokyo, Japan
Tel: [+81] 3.5463.7165