The psyche of the average Internet user includes an uncontrollable desire for instant gratification. Want something? The internet is ready and willing to provide anything to anyone able to point and click. However, this convenience comes at a cost. In most cases the cost of operating internet services are recouped via business models such as advertising, membership dues, product markup, and hidden costs. The effect of this trade off between convenience and business is of global proportions.

Cost can be presented in a myriad of ways depending on the needs of a company and the services it provides. Would a company choose a slower system to service its customers when a faster one of equal financial cost exists? When internet services are bound to a client's need of instant gratification, there is a push for service providers to acquire and deploy faster, more reliable hardware to entice additional clientele. Although the two may not always be mutually exclusive, most processors run at two speeds: hot and fast (possessing higher throughput capabilities and lower latency) or cold and slow (consumes less power)[4]. In an ideal world, if these two systems were similarly priced, providers, as they do in the real world, would more often chose the former; but the answer is not always so clear. The decision may not be as clear-cut as "performance trumps power consumption". All things considered, higher throughput and lower latency means near instantaneous consumer access, and increased power consumption is a trivial consideration when power is essentially considered to be "free", or is it?

On the most basic level, the transfer of information occurs at the bit level: the movement of electrons. An analysis [3] examining the effects of the Internet on power consumption determined that one pound of coal is required to create the necessary power for every two thousand kilobytes of data. This finding, coupled with the basic power needs of the network appliances that support the network infrastructure and the over 100 million devices connected to it, undoubtedly challenges the notion of "free power". In fact, this power demand is so enormous that it comprises thirteen percent of total U.S. power consumption. To put things in perspective, in 1999, it was estimated that the Internet was using about 290 billion kilowatt-hours per year – up from 93 billion in 1995. [3] One data center, at full load, consumes the equivalent amount of energy as 12,000 households and the electricity consumption of 38,000 households, with cooling being 30 percent of the power demand. [5] It has been proposed that when the number of PCs on the Internet reaches one billion, the amount of electricity consumption will be equivalent to the entire output of the U.S. electric grid. Intel has readily confirmed this, saying that the amount of power in a server is limited not just by the supply, but also by how much cooling is required. As a result Intel has begun initiatives to cut the power consumption of their processors such as the unveiling of a Pentium III that draws less than half a watt of power (a considerable decrease from the 33 watts a Pentium III currently draws [1]). Assume that electricity costs $0.08 per KWH. In a month, one 33 watt processor costs about 2 dollars to operate. This may seem like a small figure, but considering it costs only 3 cents to operate the half watt counterpart, data centers will certainly notice the difference in thousands of servers. Power and energy are not free and any allusion to such a statement is merely an illusion. Data centers are businesses just like any other, and would not engage in money losing practices. As the cost of energy increases, and as the amount of energy required does likewise, it is very likely that data centers will need to increase their

fees to recoup their costs, so the ultimate cost of power is returned to the data center's clients.

The cost of increased power consumption is not just limited to data centers. Service providers, may find themselves paying directly towards the increased power consumption of their machines – costs not included in the sticker price of a server. In the summer of 2000, Silicon Valley experienced a series of rolling blackouts to keep the electric grid from ceasing operations. A hundred thousand customers were affected for a period of over three days [1]. While unusually high temperatures and utility company politics contributed a fair share to the blackouts, the disproportionate number of server farms and data centers located in the bay area undoubtedly added to the high load. This problem was repeated in November of 2000, when 95 percent of all available power was consumed resulting in a "stage-two emergency" (blackouts occur more commonly when consumption surpasses 98.5 percent). What is the cost of a mere 20 minute power outage at Hewlett Packard? Where hosting companies expect more than a million dollars per minute [1] the lost time cost HP over 30 million dollars. Considering this, increased power consumption of high-speed servers could be more costly than one had previously thought. What little performance gain of an energy hogging server is quickly dwarfed when its very existence causes such costly service outages.

These hidden costs should guide industry leaders toward making strides to change the doomed course of energy consumption. It may become a necessity as utility companies become stricter, and natural resources are depleted. Companies need to begin looking at the decision of buying power efficient computers over those with a slight power advantage. Shannon Poulin, a manager with Intel said, "We get the feeling that people who buy servers aren't willing to trade away performance for lower power consumption…" Steve Cumings of Compaq agreed, "I have never heard a customer say they have made a server decision based on power consumption." [1] Fortunately, a greater number of players in the field of Internet services and application providers are slowly shifting their agendas to include discussions concerning energy efficiency. Thanks to their early attempts, it is already possible to find an acceptable level of compromise and performance. Purchasing hardware that will last a long time and finding hardware that can help to minimize the effects of high energy consumption is no longer a choice companies have to make. For example, data centers are buying into a new trend of "server blades" that combine CPU, disk drives, memory, etc in one expansion card. This combination effectively reduces the power, cooling and space each node consumes thus yielding improved reliability and lower costs [4] by only consuming a maximum of 15 watts, while their traditional server counterparts consume a relatively hefty 75-120 watts. Cooler running processors result in longer life of the components. Blade performance degradation is minimized due to the processor's ability to continuously adjust the speed to give the power/performance ratio without overheating the processor [4]. Alternative solutions such as these server blades can be custom tuned to specific applications that give companies more value for their dollar in both initial hardware cost (not just in server cost, but also in reducing the need for UPSs, generator systems, and cooling equipment [4]), as well as recurring energy costs. If the savings are not seen directly by the service

providers, it will likely trickle down from the datacenters that no longer have to support power hungry servers.

Studies analyzing the power hungry appliances of the Internet world have also been catalysts for change. Utility companies are looking to place Internet hosting companies and data centers into a different class for the purposes of rate structuring. They are gradually realizing that, at the moment, industries are not paying the true cost of service for power generation and transmission [1]. While the conclusion of one study [3] pointed to building more power plants to meet the demand, most people would agree that "more power plants" is not a scalable solution. The finite source of natural resources will see to that. A partial solution is for companies to forgo the strict "performance trumps power consumption" ideology in exchange for more flexible and sensible decision making policies that allow for non mission-critical applications to use a more power efficient machines; a step towards self preservation that could actually lead to higher availability. After all, what good is a fast transaction if a client cannot even reach the server? In fact, the bulk of availability and reliability failures do not occur at the level of the server [2], rather they are a consequence of network failures, local hosts, and operator error.

In a study commisioned in the Netherlands, trends regarding power consumption of data centers and similar such structures of the Internet were analyzed, and they proposed new energy efficient designs of data centers that reuse waste heat and optimize cooling systems that could pay back within as little as five years [5]. There is also a "coming revolution" in fuel cells and similar energy storage devices that will create system efficiencies of 90 or 92 percent [1]. Transmeta's Code Morphing Software can adjust its clock frequency according to the load thus significantly reducing wasted resources and energy [4]. The Crusoe, (developed by Transmeta), combines hardware and software in an attempt to find that elusive balance between performance and power consumption [4]. Admittedly, the Crusoe is not a powerful processor, compared to its power hungry counterparts. However, it can serve a sufficient portion of the demands of Internet services including serving web pages, e-mail and file sharing [4]. Another effortless practice by replacing Windows NT/2000 machines with Linux or deploying StrongARM processors (which peaks at a mere fifteen watts, with normal consumption at two to three watts - a 98 to 99 percent savings of electricity over Windows NT servers [1]) a plan that has the potential to not only save energy but can also save the customer money.

Currently there are many factors concerning the purchase of servers. Availability, performance, and cost play equal parts. However, many times, companies, especially those involved in the Internet infrastructure, fail to realize hidden costs when deciding which equipment to purchase. In order to reduce cost in the long term and perhaps for the greater good of the world, power consumption should become an equal factor in purchasing decisions. In many cases, customers may not even *notice* performance degradation in power efficient systems. Although technologies already exist that can greatly improve power efficiencies, and can be very effective in reducing electricity load, demand drives the market.

[1] Angel, J. “Emerging Technology: Energy Consumption and the New Economy”. Network Magazine, January 2001. http://www.networkmagazine.com/article/printableArticle?doc_id=NMG20010103S0005

[2] Merzbacher, M. and Patterson, D. “Measuring End-User Availability on the Web: Practical Experience”.

[3] Mills, M. “The Internet Begins With Coal”. October 1999.

[4]Underwood, T. “Server blades: Density versus Flexibility”. Tech Update. December 2001.

[5] Hartkamp, F. “Energy consumption of ICT”. February 2002.