A SECURE DELIVERY AND USAGE MODEL FOR RICH CONTENT THAT MINIMIZES PEAK NETWORK CONGESTION

Walter Boyles
Wavexpress, Inc.

Abstract

Broadcast is the most efficient means of delivering rich content to the consumer. Unicast pull is the most convenient method for the consumer to obtain rich content on-demand (any kind of content at any time). The inherent shared bandwidth nature of the Hybrid Fiber/Coax (HFC) network, coupled with advances in technology, allows the best attributes of both broadcast and unicast pull to be provided to the consumer with the lowest cost and a high degree of consumer choice and discretion. With this, reliable security and transactional methods must be used to protect the content and verify consumption.

Overview:

Broadcast either to the set-top-box (STB) or the cable modem can be utilized in addition to and, at times, as an alternative to on-demand delivery. For set-top box delivery, video-on-demand (VOD) allows the consumer to select, from a menu of movies, a title thus dedicating bandwidth to that consumer for the duration of the film. However the terms of sale, typically, will also allow the consumer to pause that film for up to a specified period of time (i.e. the film must be viewed within a 24-hour time period). The MSO can choose to maintain that dedicated bandwidth during the pause or reclaim it for use by other viewers. However, even if that bandwidth is reclaimed, the terms of sale frequently will mean that the MSO must be able to provide alternative bandwidth for a return to viewing in a reasonable period of time after the pause by the consumer. As such, reserve bandwidth must be allocated to make certain that that the consumer does not return from a pause and the terms of sale cannot be met by the MSO.

VOD is not yet widely deployed, but it can be anticipated that peak periods for VOD usage will exist (e.g. Saturday evenings from 7 PM – 12 AM) and that slack periods will also occur (e.g. Monday mornings from 7 AM – 11 AM). Bandwidth utilization for VOD will be extremely uneven.

For cable modem traffic, peak periods of utilization may similarly occur in evenings when consumers will experience noticeable degradations in performance due to bandwidth demand within the HFC network and/or to delays outside of the network (e.g. many hits on a server) as compared to early mornings when demand within the network as well as overall Internet traffic may be less. In either case, a way to equalize loading on the network over time provides benefits to consumers and to operators alike.

Storage Technology:

Over the last ten years, while Moore's law has become a well-known and widely quoted phenomenon, the time duration between the doubling of chip densities (the number of transistors on a silicon substrate), has been slowly increasing to upwards of 18 months.

Meanwhile, advances in magnetics technology have allowed for the doubling of hard drive densities in half that time (approximately nine months). During the same period, while the cost of hard drive

storage has been decreasing rapidly, the last mile bandwidth available to the average consumer has been increasing relatively slowly. Even for those fortunate consumers with broadband connectivity, the price and the speed of their connection has changed little over the past few years.

Delivering content in broadcast or multicast modes, that is subsequently cached, can allow consumers faster access to that content. This type of delivery is, in part, enabled by large amounts of inexpensive storage being made available to the consumer.

Year	1988	1998	1999	2002 (predicted)
Cost/ MB	$11.54	$0.04	$0.02	$0.003

Table 1. Hard Drive Density vs. Cost (Disk/Trend).

Table 1. shows industry mean storage prices. However, consumers buying desktop PCs typically buy in the most economical segment of the hard drive industry. Today for mid-range desktop systems, the cost of storage is often already at $0.003 per incremental MB.

Additionally, with drive densities doubling every 9 months, the cost of storage in the lowest cost part of the curve, in early 2004, may be extrapolated to be as low as one hundredth of a cent per MB. That calculates to a cost of approximately 35 cents to store 3.5 GB video:

$$3500 \text{ MB} \times \$0.0001 /\text{MB} = \$ 0.35$$

Without massive increases in last mile bandwidth to the consumer, the trend will be that it is cheaper to store large files then to transmit them in unicast.

And with drives approaching 200 GB already available and mid-range desktop systems typically equipped with 20 - 80 GB drives, storage already greatly exceeds the space taken up by the operating system and typical applications software packages. Further, again extrapolating drive densities forward we can expect that by the end of 2002, drives exceeding 1,000 GB (1 Terabyte) will be available and new mid-range consumer systems should nominally have hard drives of 100 - 400 GB. This reduction in the price of storage will equally allow the addition of large amounts of storage to be added to set-top boxes.

Now, as with Moore's law and the increase of densities of integrated circuits, it is well recognized that increases in drive density eventually run into barriers imposed by the laws of physics. In the case of magnetics, one barrier is the supermagentic effect (SPE). The SPE limitation will occur when the amount of energy of the magnetic spin in the atoms, that constitutes one bit of information, approaches the ambient thermal energy. When this occurs, the bits become subject to random "flipping". It is believed that the limitations imposed by SPE may limit miniaturization as early as 2005 at around 150 Gb per square inch.

Beyond this 150 Gb per square inch limit, new strategies are already being devised. These include techniques from changing the orientation of bits to new magnetic materials and even to moving towards the addition of optical materials to magnetic materials. With these kinds of techniques it is entirely possible to conceive of consumers possessing many terabytes of storage on their PCs in the not so distant future.

With the amounts of storage becoming available in the shorter term, the delivery of major interest movies to all subscribers on a network (who elect receipt) will be possible while allowing the same viewing flexibility as VOD with a single uninterrupted transmission by the MSO. Similarly, one could allocate a portion of the 38 Mb/s bandwidth (based on

256 QAM) of a 6 MHz cable modem channel to broadcast during off-peak hours. Delivering 5Mb/s during the 2 AM- 7 AM period of low-usage would mean the availability of over 100 GB of content to the cable modem customer:

(5 Mb/s x 3600s/hr x 5 hr) / 8 = 112.5GB

The consumer would have the ability to elect to cache any portion or all of that day's delivery of content.

This would allow the consumer to spend a portion of his viewing time (most likely in the peak usage period of the evening) consuming that cached content and thus further limiting peak usage. Thus, the use of caching content received from off-peak delivery can result in the usage of excess bandwidth in slack hours and decreased demand during peak hours. The bandwidth available and bandwidth demand are then more equalized and the network operates closer to maximum efficiency.

Security:

In delivery of stored content, a number of security mechanisms become extremely important. First, a reliable system of copy protection for content cached to the drive of the PC or the STB is needed. Second, a security system that allows protection of royalties for content consumed is required. Third, a reliable and cost-effective method of transacting on the content purchased, leased or viewed is a key element.

To satisfy these requirements, the security of the content, the permission to access that content, and even the transaction on the content itself are most reliably achieved through a hardware security device. This conclusion is consistent with the methods used today by the cable industry in allowing access to premium content where security is contained within the STB. Monolithic solutions are, of course, preferred that do not allow hackers access to any bus containing cleartext data.

The rationale for hardware security involves many factors but to a degree it comes down largely to one. Software is relatively easy to tamper with and software hacks are easy to replicate on a large scale. The delivery and consumption of content is far different from an Internet purchase of a physical good where security involves encrypting a credit card number for a purchase that the consumer is authorizing. In content security, there is a temptation and an opportunity to avoid paying for something in an area where precedents like unprotected MP3s already exist.

A hardware device that allows some degree of programmability and renewability is also preferred. First, a programmable device can provide multiple security functions including persistent protection of content which is vitally important in an open system like a PC. Second, a system with some degree of renewability can utilize a smart card or preferably, for true hardware renewability, a more complex portable token that avoids some of the limitations (e.g. interface and performance) of an ISO 7816 smart card. If smart cards are used, care must be taken to avoid unsecured smart card readers in a PC environment since key security inside the PC is essential to maintaining content security.

The requirements for security in a PC for copy protection are different than in consumer electronic devices. In consumer electronic devices, a bus-oriented copy protection system, between say a DVD player and a monitor, may appear adequate. However, in a PC, the variety of ways to access that content inside the PC plus the number of ports make content protection a more complex issue than in the consumer electronics space.

Digital rights management systems (DRMs) have proliferated in the PC space although their actual usage has been somewhat limited due, in large part, to consumer behavior and lack of consumer

acceptance. However, software DRMs still exhibit the same security weaknesses as other software solutions do and provide less resistance to tampering than well thought out hardware solutions including hardened DRM.

Additionally, since content delivered and cached will typically be viewed on a time-shifted basis, the ability to securely transact this content will be also be of importance. Ideally, a set of security mechanisms and methods for processing transactions off-line would provide the benefit of not requiring repetitive server transactions. Content consumed and paid for, in this way, would provide the benefit of minimizing operator cost in the same way as the single delivery of a copy of content to multiple consumers in the network did. The decryption and transaction are most reliably secured in a single tamperproof hardware device attached to the PC or embedded in the set-top box.

Since the cable modem-attached computer receives content over multicast and frequently the interactive STB has a back channel available, a strong key exchange method is required. The security mechanisms (including copy protection) should avoid use of hidden global secrets or other mechanisms, which may result in catastrophic failures.

Users and preferably legitimate devices (both) should be authenticated by a reliable process that disallows common known attacks. Authenticated users should be able to prove that they possess legitimate devices for metering and decryption of content.

Unique IDs are widely used for to identify devices such as the IEEE 802.3 48 bit addresses in NIC cards. However, for maximum security, IDs should not only be immutable (not able to be reprogrammed by the hacker) but shielded (secret) from the consumer. Additionally, legitimate hardware IDs should not be able to be guessed by potential hackers.

One mechanism for identifying legitimate devices is the assignment of non-deterministically, random identities at the time of manufacture of devices. By assigning immutable secret identities, of adequate length to devices, illicit clones without legitimately assigned identities are detected. The attacker's ability to guess one legitimately assigned random number represents a problem comparable to exhaustive key search for a strong symmetric key cryptographic algorithm. One issue is the potential for the occurrence of multiple instances of any one legitimately assigned random ID.

For example, if a 128 bit random identity is assigned, this constitutes a total of 3.4×10^{38} different random numbers. However, if two hundred million consumer devices (2×10^{8}) devices are built the probability that no two of these devices possess the same random number is:

$$\prod_{n=1}^{n=x-1} \left(1 - \frac{1}{y}\right).$$

Where x = the number of devices and y= the number of possible random numbers. Therefore, the probability that there are multiple instances of one ID is:

$$p = \left(1 - \prod_{n=1}^{n=199,999,999} \left(1 - \frac{n}{3.4 \times 10^{38}}\right)\right) > 0.$$

Thus, there exists a possibility of multiple instances of the same ID and one cannot automatically assume the existence of multiple instances constitutes an illegal clone.

The binding of device identities to user identities (the user identities may be kept in a smart card or other portable token) can be

used as a barrier to useful theft of the device. However, the value of having a user identity bound with a random device identity at the time of purchase also provides a more effective barrier to random guesses. That is, the attacker (guesser) must not only guess an assigned (secret) device identity but also know the user identity to which it is bound. Many guesses with the same user identity are easily detected. Also, since there is no longer sole reliance on device IDs, the problem of multiple instances of the same legitimately assigned device identity can be minimized.

Once a legitimate device and user are determined, a user's security device can participate in a public key exchange and the sending of the symmetric keys to allow the access to content. In some cases, the keys transmitted will not decrypt the content itself but rather other keys that in turn are used to decrypt the content. Changing keys is done on a periodic basis; in some cases, systems require that many key changes be made in a relatively short period of time.

The decryption of the cached content should occur subsequently or coincide with a payment for that content. Payments can occur either on-line or off-line. Secure off-line transactions have the advantage of minimizing server usage and often also transaction fees. Off-line transactions, of this nature, are typically debits and additional security, for the transaction being recorded, can be provided by a real-time clock in the security/debit mechanism.

Once content is transacted on, a method of usage must be employed to prevent unauthorized redistribution of content. In a closed system, a local usage method may be employed or instead one where a smart card accompanies the content with a consumer to other devices/players. In an open system such as a PC, persistent protection by some system such as digital rights management technology may be employed.

The use of a programmable security processor can allow purchase, decryption, and persistent protection inside the same device. Devices of this nature must have a well-conceived trust assurance network for loading code. A secure operating system within these devices may also be desirable.

A programmable security processor will contain secure volatile and nonvolatile memory. The limitation of a monolithic solution containing volatile and nonvolatile (i.e. for key storage) memory is often the amount of volatile memory (typically SRAM) that can be placed economically on-chip. This limitation is due to the processing difficulties of fabricating custom ICs with both DRAM and nonvolatile memory (e.g. Flash or E^2PROM).

In the future, devices of this nature may be fabricated with ferroelectric memory (FRAM) that may take the place of both the volatile and nonvolatile memory blocks thus allowing massive code spaces on-chip. This type of chip architecture would allow very complex content usage and rights management software to be loaded entirely inside a single chip thus maximizing security.

Conclusion:

The cable industry has continued to innovate products and services to the consumer. For consumers, the movement towards cable modems and digital cable, is enabling access to much more content with a greater degree of ease than ever before.

Movement to more efficient modulation and compression techniques coupled with advances in architectures allows MSOs to more effectively use the capacity of their system. However, the addition of delivering content during off-peak hours in broadcast to the STB or IP multicast to the cable modem

consumer's hard drive enable even greater efficiencies by the MSO and more choice and greater access to the consumer.

The underlying technologies, that allow business models which employ off-peak delivery and caching of encrypted content for subsequent consumption and off-line payment, are increases in storage, strong security mechanisms including copy protection, and reliable methods of transacting on the content that is to be consumed.

Bibliography:

Data–Over-Cable Service Interface Specification, Cable Modem to Customer Premise Equipment Interface Specification, Cable Television Labs, 2000.

Dell, http://www.dellforme.com, 2001.

Huitema C., IPv6, Prentice Hall, 1997.

Toigo, Jon W., Avoiding a Data Crunch, Scientific American, May 2000.