MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
IEEE COMMUNICATIONS SOCIETY
http://www.comsoc.org/~mmc

E-LETTER

Vol. 6, No. 5, May 2011

IEEE COMMUNICATIONS SOCIETY

CONTENTS

MESSAGE FROM MMTC CHAIR

Dear MMTC fellow members,

At the beginning, I would like to salute to the Japanese people, who have shown extreme calm and stoicism during the recent damage, although their lives were affected by the earthquake, tsunami and nuclear crisis. We hope the situation will get recovered soon to get people back into the normal life. Apparently many of our members' travel plans of attending ICC'11 at Kyoto were affected also, thus we decided to cancel our TC meet originally planned to be held in ICC'11 and move the meeting to ICME'11 to be held in July at Barcelona, Spain.

In the past weeks, the Award Board led by Drs. Mung Chiang (Princeton University, USA) and Song Ci (University of Nebraska-Lincoln, USA) have been working very hard to select the winners of a number of paper and service Awards from the many nominees. Now I take the privilege to announce the final outcome:

(1) Year 2011 Outstanding Leadership Award

Winners:

Lingfen Sun, Univ. Plymouth, UK
(Chair, QoE Interest Group, MMTC)

Fen Hou, Macau Polytechnic Institute, China
(Co-Director, Membership Board, MMTC)

(2) Year 2011 Best Journal Paper Award

A. Dua, C. W. Chan, N. Bambos, and J. Apostolopoulos, "Channel, deadline, and distortion (CD2) aware scheduling for video streams over wireless", *IEEE Transactions on Wireless Communications*, vol. 9, no. 3, pp. 1001-1011, March 2010.

Authors:

Aditya Dua, Qualcomm, USA
Carri Chan, Columbia University, USA
Nicholas Bambos, Stanford Univ., USA
John Apostolopoulos, HP Labs, USA

(3) Year 2011 Best Conference Paper Award

S. Gunawardena and W. Zhuang, "Voice capacity of cognitive radio networks", *Proc. IEEE ICC*, May 2010.

Authors:

Subodha Gunawardena, Univ. of Waterloo, Canada
Weihua Zhuang, Univ. of Waterloo, Canada

Let us congratulate all Award winners and hope they continue their endeavor to make new contributions to our community. In addition, we would like to thank the Award Board members, and members who made nominations to us. It is worth to mention our policy for Best Paper Award again here, in order to get your paper qualified for a Best Paper Award, please nominate it to our R-Letter Board first, as all papers reviewed in the R-Letter will automatically enter the pool of Best Paper Award nominees.

At last but not the least, I would like to encourage our members to attend our TC meeting at ICME 2011 (Barcelona, Spain). More information of this event can be found at ICME official website: http://www.icme2011.org/

Thank you very much!

Haohong Wang
Chair of Multimedia Communication TC of IEEE ComSoc

SPECIAL ISSUE ON DECISION AND GAME THEORY FOR SECURITY

Decision and Game Theory for Security (Part 2)

Jiangtao Wen, Tsinghua University, Beijing, China
Tansu Alpcan , Technical University of Berlin, Germany
jtwen@tsinghua.edu.cn, alpcan@sec.t-labs.tu-berlin.de

The second part of the special issue on decision and game theory for security explores the frontier of incorporating game theory into the research of security, especially multimedia security, in a highly networked world where access to multimedia content anywhere, any time, on any devices is increasingly required and taken for granted. The distributed manner in which the content is processed and transmitted, as well as the locations where an attack could be launched might be naturally modeled by a multi-party competitive game.

The networked nature of multimedia production, transmission and consumption poses unique challenges to security, including

- The vast differences in the devices that are used to produce and consume multimedia content, including both processing power and security/temper resistance of the system
- The need for, in most cases lossy, transcoding/transrating of the content during transmission
- The impact of network losses on content encryption and watermarking
- Heightened diversity and rapid change in content value over the content life time, especially for user generated content
- Large diversity of the sophistication of content producers, aggregators and distributors
- New protocols and paradigms such as peer to peer for distributing content.

The first two papers in this special issue were originally presented at GameSec 2010, the inaugural Conference on Decision and Game Theory for Security hosted the Technical University Berlin, Germany, on November 22-23, 2010.

The article “Security Investment with Penetration Testing” is concerned with the critical issue of where to allocate resources when design a security system, by way of optimizing the return on security investment using applying the Iterative Weakest Link (IWL) approach for the analysis of the impact of penetration testing.

In “Negative Externalities from Weak Password Practices on the Web”, the authors provided insights into the challenges of managing login credentials, especially username/passwords in an 100-account world, namely, the problem for the user to “map” appropriately cryptographically strong credentials to sites with the corresponding loss profile, and the challenge of incentivizing the web sites to design the appropriate authentication/security system so that the design not only meets the need for the site itself, but also behaves as a friendly citizen in a networked world where the same user/password may be used to authenticate the user for a large number of sites with different security requirements and capabilities.

The lasts two papers in this issue further illustrate application of game theory to multimedia security. In the last decade, the technology and adoption of peer to peer (P2P) based content distribution systems have made significant strides. P2P systems are widely used in popular services such as Skype or even video conferencing and messaging on mobile terminals. The article “A Game-Theoretic Interpretation of P2P Swarm Equilibrium for Incentive Optimization” provides a game-theoretic analysis of the popular BitTorrent protocol and shows that by designing a new protocol (prTorrent) that establishes a Nash Equilibrium within the swarm of the system, both the overall security and performance (e.g. bandwidth utility) of the system can be significantly improved.

Finally, finger-printing techniques are widely used in many content authentication and piracy tracing systems, and are becoming ever more important when tens of millions of hours of user generated content (UGC) are produced and consumed every year, while at the same time, content distribution, copying and editing have become increasingly cheaper. It is critical to understand the dynamics between an attacker of a finger-printing system and its designer, so as to provide important heuristics and insights that could be used to improve the security of the overall system in a highly diversified environment.

The article "Security Investment with Penetration Testing" is concerned with the critical issue of where to allocate resources when designing a security system. It focuses on penetration testing, a common and specific practice to information security, whereby commissioned testers penetrate and investigate the target system from an attacker's point of view, and these testers report weaknesses rather than exploit them. The article discusses resource allocation for security system design by way of optimizing the return on security investment using the Iterative Weakest Link (IWL) approach to analyzing the impact of penetration testing.

Jiangtao (Gene) Wen received the BS, MS and Ph.D. degrees with honors from Tsinghua University, Beijing, China, in 1992, 1994 and 1996 respectively, all in Electrical Engineering. From 1996 to 1998, he was a Staff Research Fellow at UCLA, where he conducted research on multimedia coding and communications. Many of his inventions were later adopted by international standards such as H.263, MPEG and H.264. After UCLA, he served as the Principal Scientist of PacketVideo Corp., the CTO of Morphbius Technology Inc., the Director of Video Codec Technologies of Mobilygen Corp, the Senior Director of Technology of Ortiva Wireless and consulted for Stretch Inc. and Ocarina Networks. Since 2009, Dr. Wen has held a Professorship at the Department of Computer Science and Technology of Tsinghua University. He was a Visiting Fellow at Princeton University in 2010. Dr. Wen's research focuses on multimedia communication over challenging networks and computational photography. He has authored many widely referenced papers in related fields. Products deploying technologies that Dr. Wen developed are currently widely used worldwide. Dr. Wen holds over 30 patents with numerous others pending. A Senior Member of IEEE, Dr. Wen is an Associate Editor for IEEE Transactions CSVT.

Tansu Alpcan received the B.S. degree in electrical engineering from Bogazici University, Istanbul, Turkey in 1998. He received the M.S. and Ph.D. degrees in electrical and computer engineering from University of Illinois at Urbana-Champaign (UIUC) in 2001 and 2006, respectively. His research involves applications of distributed decision making, game theory, and control to various security and resource allocation problems in complex and networked systems. He is recipient of multiple research and best paper awards from UIUC and IEEE. He has taken part in organization of several workshops and conferences such as IEEE Infocom, GameComm, and GameSec as TPC member, associate editor, co-chair, chair, and steering board member. He is the (co-)author of more than 100 journal and conference articles, an edited volume, as well as the book "Network Security: A Decision and Game Theoretic Approach" published by Cambridge University Press in 2011. He has worked as a senior research scientist in Deutsche Telekom Laboratories, Berlin, Germany, between 2006 and 2009. Tansu Alpcan is currently assistant professor (Juniorprofessur) in Technical University Berlin while continuing his affiliation with Deutsche Telekom Laboratories.

Security Investment with Penetration Testing

Rainer Böhme, University of Münster / ERCIS, Germany
rainer.boehme@wi.uni-muenster.de
Márk Félegyházi, Budapest University of Technology and Economics, Hungary
mfelegyhazi@crysys.hu

1. Information Security Investment

In the past couple of years, information security investment decisions have increasingly attracted the attention of scholars in computer science, economics, management science, and related disciplines. Works in the early 2000s were mainly concerned about *how much* security investment is enough, using modeling approaches borrowed from general quantitative investment theory. More recent research has shifted the focus to more domain-specific questions, such as *where to invest*. In other words, the helicopter perspective of defining a security budget has been replaced by a strategic view on allocating security investments between different investment alternatives, such as perimeter versus internal access control, proactive versus reactive defense, awareness raising versus new equipment, and so forth. Obviously, this more detailed view requires modeling approaches specific to the security domain and thus makes it a more interesting research area for security engineers. [2]

Penetration testing (short: pentesting) is a practice that is both common in and very specific to the information security domain with hardly any analogies in other industries. Penetration testing is also referred to as “ethical hacking” because the commissioned penetration testers investigate the target system from an attacker's point of view, reporting weaknesses rather than exploiting them [1]. It is widely used in practice, but its effects have not been reflected in the information security investment literature. Our contribution in [3] is to extend a domain-specific model of security investment by incorporating the option to commission pentests. Because pentesting reduces the defender’s uncertainty about possible threats, it is convenient to build on a model that explains defense strategies with uncertainty reduction. Hence we chose the “iterated weakest link” (IWL) model [4].

2. The Iterated Weakest Link (IWL)

The original IWL model explains why a defender facing uncertainty about which threats are most likely to realize might defer security investment and learn from observed attacks where the investment is most needed. The benefits of more targeted investment may outweigh the losses suffered through non-catastrophic attacks, thereby increasing the *return on security investment* (ROSI).

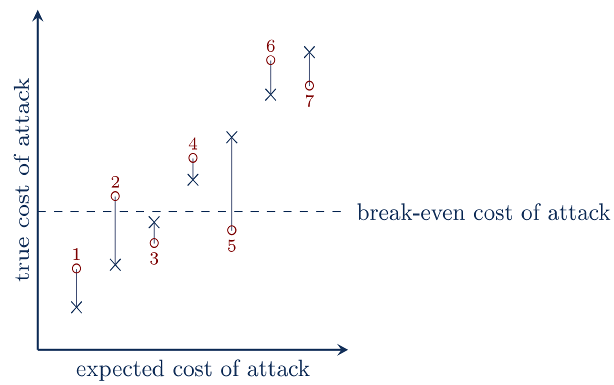


Fig. 1. Defender’s uncertainty in the IWL model

The IWL assumes that the defender can enumerate a list of possible threats. She further has some intuition of their expected difficulty to attack expressed by a sequence of threats ordered by increasing cost of attack. These expected costs of attack are to be distinguished from the true costs of attack unknown to the defender. In the example of Fig. 1, a defender who is bound to *proactive* defenses would invest in protection against the most likely threats (1, 2, 3) below the break-even cost of the attacker only to find that she remains vulnerable through thread 5. Now if losses due to attacks are not catastrophic and defenses can be upgraded *reactively*, then losses due to attack can be accounted as costs in a finite-horizon discrete-time player-versus-nature game. A superior strategy in the given example is to protect only against threat 1 proactively, and then wait for the attacks to take place. Successful attacks reveal which links are weaker than expected. Hence defenses against threats 3 and 5 should be added reactively in two subsequent rounds resulting in the least costly secure defense configuration (1,3,5) after two rounds.

The left column of Tab. 1 qualitatively reports the advantages of a combined reactive defense strategy over a purely proactive strategy. Security spending drops, but every dollar is spent more efficiently as witnessed by an increase in ROSI. A side-effect of successful attacks being tolerated to learn unknown true costs of attack is that the overall attack intensity raises.

Tab. 1. Effect of investment in penetration testing

	Reactive defense (vs proactive)	**Pentesting** (vs reactive w/o pentesting)
Attack intensity	↗	↘
Security spending	↘	↘
ROSI	↗	↗
ROPT		↗

3. Decision to Hire Penetration Testers

We extend the IWL model to study the impact of penetration testing. Penetration testing is included in the model as an option to gather information prior to investing into protection against so-identified threats. Pentesting extends the action space of the defender in every round. Our intuition is that pentests, just like attacks, reveal information about the next weakest link, thereby reducing uncertainty. In this sense, actual attacks and pentests are substitutes.

When comparing reactive defense *with* pentesting to reactive defense *without* pentesting (right column of Tab. 1) in our extended model, security spending drops further. This is because the option to commission pentests allows the defender to start at lower optimal levels of proactive defense. This reduces the probability of misallocation. Consequently, ROSI improves and a novel metric called *return on penetration testing (ROPT)* is strictly positive. Using this metric, we show that pentesting, if reasonably priced, brings a substantial benefit by providing advanced intelligence information to the defender. As an additional benefit, the attack intensity against the defended systems decreases, which is relevant in the presence of externalities or indirect costs, such as reputation loss.

4. Conclusions

Many research papers study optimal security investments and they mostly focus on how much a defender has to invest to protect her system. In our paper [3], we investigate the *investment options* of the defender. More precisely, we build on the IWL model of security investment under uncertainty and show that information-gathering options, such as penetration testing, bring a significant benefit to the defender.

Directions for further research include empirical validations of some of the model's implications as well as a relaxation of selected modeling assumptions. For example, the model can be reformulated for an infinite horizon in continuous time with discounting.

Our analyses emphasize that the results of penetration testing should be considered in the light of risk assessment rather than perceived merely as a security checklist. This holds for fairly general parameter choices. Conversely, providers of penetration testing can use our model to price their services based on the defenders' willingness to pay rather than costs plus markup.

References

[1] Brad Akin, Scott Stender and Gary McGraw: Software Penetration Testing. *IEEE Security & Privacy*, **3** (1), pages 84–87, 2005.
[2] Rainer Böhme: Security Metrics and Security Investment Models. In I. Echizen, N. Kunihiro and R. Sasaki (eds.): Advances in Information and Computer Security (IWSEC '10), pages 10–24, LNCS 6434, Berlin Heidelberg, 2010. Springer.
[3] Rainer Böhme and Márk Félegyházi: Optimal Information Security Investment with Penetration Testing. In T. Alpcan, L. Buttyán and J. S. Baras (eds.): Decision and Game Theory for Security (GameSec '10), pages 21–37, LNCS 6442, Berlin Heidelberg, 2010. Springer.
[4] Rainer Böhme and Tyler Moore: The Iterated Weakest Link. *IEEE Security & Privacy*, **8** (1), pages 53–55, 2010.

Rainer Böhme is Assistant Professor of Information Systems and IT Security at the European Research Center for Information Systems (ERCIS), of the University of Münster in Germany. Prior to that he was Postdoctoral Fellow at the International Computer Science Institute in Berkeley, California.

His research interests include economics of information security and privacy, privacy-enhancing technologies, multimedia security, and digital forensics. He holds a Master's degree in Communication Science and Economics and a Doctorate in Computer Science, both from Technische Universität Dresden in Germany. He also worked at the European Central Bank, where he served for the directorates of economics and financial stability & supervision.

Márk Félegyházi is Assistant Professor in the Laboratory of Cryptography and Systems Security (CrySyS) at the Budapest University of Technology and Economics in Hungary. Prior to that he was Postdoctoral Fellow at UC Berkeley and the International Computer Science Institute in Berkeley, California.

His research interests include game theory in wireless and ad hoc networks, information security policy, and measurement aspects of Internet attacks and defense. He holds a joint Diploma degree in Electrical Engineering from Budapest University of Technology and Economics, Hungary, and Universität Karlsruhe, Germany. He obtained his PhD in Communication Systems from EPFL Lausanne, Switzerland. He also worked with Ericsson Research in Hungary.

Negative externalities from weak password practices on the Web

Sören Preibusch and Joseph Bonneau, University of Cambridge, Cambridge, UK
{sdp36, jcb82}@cl.cam.ac.uk

1. Incentives behind password ubiquity

The combination of username and password is used ubiquitously on the Web to authenticate users. The average Web user is now thought to use 25 passwords for personal Web browsing (Florêncio & Herley, 2007), and our own data indicates that more than 90% of the 100 most visited destinations on the Web and more than 80% of the top 1000 Web sites collect passwords from their users. Neither cryptographic enhancements to user authentication nor password replacing single sign-on schemes have seen wide adoption yet.

The incentives for setting up password schemes are twofold. First, they enable cheap authentication, by making system access conditional on the user knowing the correct credentials. Merchant Web sites, which often enable users to store payment details for future shopping, have direct incentives to secure password authentication, at least to evade liability. Password-secured user accounts are also intended to make users accountable for their actions and eliminate plausible deniability (e.g., online forums, social networking sites, Web mail).

The intrinsic motivation to secure password installations is missing for Web site operators that use username-password combinations for content customisation. In particular, news sites use accounts to personalise their content for registered users. It is often unclear why such services require authentication. As long as users can create an unlimited number of free accounts password authentication seems dispensable. However, the personal information collected during account creation, including socio-demographic details, may be valuable to the Web site operator for behavioural advertising.

Regardless of the motives, there are costs associated with deploying password authentication which rise with the sophistication of the mechanism. These include at least the need for skilled programmers to implement and maintain the system and the operating costs of password storage.

2. Negative exteranlities in password deployment

The benefits for each company from deploying a password scheme come at the expense of the community, a phenomenon known as negative externality.

Given the low adoption of password management tools, users' mental storage capacity for passwords is limited. As Web companies are not charged for imposing the burden of remembering another password, users' memory is overused in a tragedy of the commons. Users often respond by reusing passwords across sites.

Password reuse introduces a further negative externality in the password market. Companies' differing motives to deploy password schemes—outlined above—result in varying levels of password security and care by site operators. Because low-security sites do not bear the social costs of weak password practices, there is a tendency for under-investment in good password practices. The accumulation of high-security login credentials at Web sites with weaker incentives for security presents an attractive attack strategy: compromise a low-security site and attempt to use the login credentials at higher security sites. Such attacks have already occurred (e.g. RockYou, 2010).

3. Game-theoretic model

We have developed a game-theoretic model with low-security and high-security Web sites as players to analyse incentives for investing in good password practices (Preibusch & Bonneau, 2010).

We take into account the costs of deploying password schemes of varying sophistication, plus the benefits derived from collecting personal information for low-security Web sites and keeping payment systems secure at concerned Web sites. Interaction effects are captured by negatively affecting the payoff of high-security Web sites if password reuse occurs amongst the shared user base.

The password game has two robust equilibria. First, the security-indifferent Web site accepts any password, and the security-concerned Web site requires strong passwords (in rare cases, it may allow weak passwords as well, depending on the relative importance of extra password strength compared to costs of assessing and enforcing password complexity). The second group of equilibria is reached if the security-indifferent Web site renounces password collection. Again, the concerned Web site may enforce strong passwords or accept weak and strong.

The first group of Nash-equilibria is inefficient due to negative externalities originating in password reuse. Collectively, absence of password collection at security-indifferent Web sites is better (Pareto-superior). This social optimum is typically not reached due to incentives for password deployments. However, the game-theoretic model reveals that security-concerned Web sites should be willing to subsidise other Web sites to make them renounce password collection or provide technical assistance in deploying weak credentials only. The details of the analysis are reported elsewhere (Preibusch & Bonneau, 2010).

4. Password deployments: empirical evidence

We challenged the predicted market outcome of the game-theoretic model by two large-scale assessments of password implementations in the wild.

A manual analysis of the password practices at 150 public Web sites in the areas of electronic commerce, identity services (e.g. emailing and social networking), and content services such as news, reveals universal technical and policy shortcomings. With high significance, however, password carelessness is more prevalent at news sites, whose business is usually security-indifferent. They are also significantly more likely to collect personal information at the time of password creation. All but 2% collect email addresses on a mandatory basis and they are more likely to verify these with very high significance.

Security enhancing mechanisms such as encrypted transmission, limits on guessing attacks, and advice for users are significantly less prevalent at news sites. Conversely, Web sites with merchant facilities are significantly more likely to impose minimum password lengths and to blacklist common passwords. By interpreting merchant Web sites as security-concerned operators, these results directly support the predictions of the password game. (Bonneau & Preibusch, 2010)

5. Prevention of password sharing: empirical evidence

In a second experiment, we measured Web site operators' password care by their eagerness to prevent password sharing amongst their users. Using portals such as BugMeNot, Web users can swap credentials "to quickly bypass the login of Web sites that require compulsory registration and/or the collection of personal/demographic information" (www.bugmenot.com). Users of BugMeNot can upload new username/password combinations for a given Web site, which other visitors of BugMeNot can then retrieve for free. Web site operators can request to be removed from BugMeNot, and it is possible to check which sites have taken this step (Figure 2).

We matched the 3172 most popular US Web sites from the Alexa Top Sites ranking against the BugMeNot database and recorded for each match whether the site was blocked, missing or present. For each Web site, we also checked whether it was a news site, using a combination of manual coding by two raters and a semi-automated match against the Web site classification from the Open Directory Project.

In total, 988 of the top sites were not listed in BugMeNot; of the remaining 2,184 sites, 531 were "blocked" while 1,653 had credentials listed.

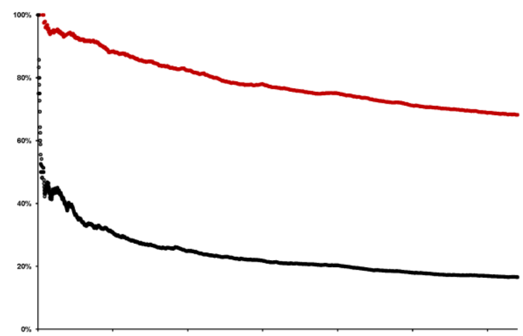


Figure 1: Proportion of top US Web sites that collect passwords (red/upper line) and of Web sites that block password sharing (black/lower)

Very-high traffic Web sites are much more likely to block listing of their credentials on BugMeNot (Figure 1; 28% blocked in upper half compared to 20% in lower half, highly significant). However, blocking remains common among the lower ranked sites, at above 17% up to rank 3000. Password collection remains at a high level of above 80% through the entire range up to the 1000th rank and is still at 69% for up to the 3000th rank. The BugMeNot blocking data is therefore a useful indicator of Web sites' real security motivations for collecting passwords.

News sites are very significantly less likely to block credential sharing ($p < 0.0001$ using a two-tailed G-test): 26% of the non-news Web sites on

BugMeNot are blocked compared to only 9% of the news sites. Thus, the BugMeNot data provides strong evidence that news Web sites have lower security concerns than other password-collecting Web sites.

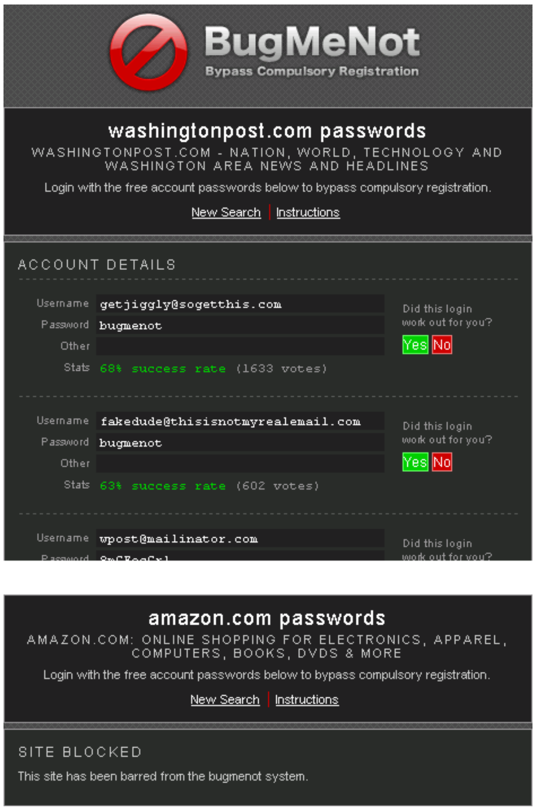


Figure 2: Screenshots of the BugMeNot service: multiple username/password combinations are shared for the news Web site washingtonpost.com, but the merchant Web site Amazon is blocked.

6. Conclusion

Password-equipped Web sites fall into two broad categories: sites which have a self-interest to invest in security and sites which primarily use passwords as a trigger to collect personal information from their users. Password carelessness is significantly more prevalent amongst the latter. The resulting dichotomy in the market corresponds to the prediction of a game-theoretical model where security-indifferent and security-sensitive Web site operators make password deployment choices. Their differing motivations lead to differing optimal choices as to what combination of weak and strong passwords should be accepted, bearing in mind the required implementation efforts as well as increased levels of protection.

Owing to negative security externalities from weak passwords, the allocation of password strength is inefficient on the market. This calls for a combination of regulation and technical approaches to pricing password deployments (Preibusch & Bonneau, 2010).

References

[1] Joseph Bonneau and Sören Preibusch, "The password thicket: technical and market failures in human authentication on the web," in Ninth Workshop on the Economics of Information Security (WEIS 2010), Harvard, 2010.

[2] Dinei Florêncio and Cormac Herley, "A large-scale study of web password habits," in WWW '07: Proceedings of the 16th international conference on World Wide Web, New York, 2007, pp. 657-666.

[3] Sören Preibusch and Joseph Bonneau, "The Password Game: Negative Externalities from Weak Password Practices," in Decision and Game Theory for Security (GameSec), vol. 6442 (Lecture Notes in Computer Science), Berlin, 2010, pp. 192-207.

Sören Preibusch is completing his PhD on negotiable privacy policies at the Computer Laboratory, University of Cambridge. He holds a diploma in industrial engineering from Technical University Berlin and has been a scholar of the German National Academic Foundation since 2003.

Joseph Bonneau is a PhD candidate in the Security Group at the Computer Laboratory, University of Cambridge. He is a Gates Cambridge Scholar. He holds BS and MS degrees from Stanford University.

Game-Theoretic Modeling of Video Fingerprinting

Avinash L. Varna and Min Wu, University of Maryland, College Park, USA
{varna, minwu}@umd.edu

1. Content Fingerprints

Multimedia consumption via the Internet has increased radically over the last few years. Multimedia content such as video and audio are frequently distributed through the internet. Video streaming services are available from such providers as Netflix, Blockbuster, Hulu, and Amazon. Services such as Google TV and Apple TV that are being planned will further strengthen this trend. Fueling this trend is the technological improvement in the bandwidth of network connections, and the growing popularity of user-generated content (UGC) websites such as YouTube. UGCs have changed the perspectives of both content providers and consumers with regards to the internet.

At the same time, the popularity of UGC websites has raised a challenge in protecting intellectual property rights, as copyrighted content may be uploaded to these websites by users. Service providers would like to identify copyrighted content uploaded to their websites, and distinguish it from non-copyrighted content. Automated video identification using "*digital fingerprints*" is a promising solution to implementing such content filtering technologies.

Content fingerprints are compact representations of robust and distinct properties of multimedia that can be used for identification. In this sense, multimedia fingerprints are similar to human fingerprints, and hence the name. Various techniques for constructing such fingerprints have been proposed in the literature, a review of which may be found in [1].

Recently, theoretical modeling and analysis of content fingerprinting techniques has attracted attention [2-5]. The research has focused on analyzing various modules employed in content fingerprinting algorithms, and understanding how the performance would scale when the algorithm is used for identification over large databases containing millions of video. Theoretical analysis can also provide guidelines for designing better fingerprinting algorithms.

One particular aspect of interest is the interaction between the fingerprint designer and an adversary who seeks to evade identification of the content to be uploaded. The conflicting objectives of these two parties can be analyzed under a game-theoretic framework to identify optimal strategies. In this articler, we illustrate this game-theoretic approach to modeling the content identification problem using the example of independent and identically distributed (i.i.d.) binary fingerprints, which are commonly used in identification applications [5,6]. Through this analysis we provide a firm foundation for the intuition that fingerprints with bits that are equally likely to be 0 or 1 is beneficial for the fingerprint designer.

2. System Model

We model content fingerprinting as a two-player game between the fingerprint designer D and the adversary A, who is trying to upload copyrighted content while evading detection. In this example of i.i.d. binary fingerprints, the designer chooses the distribution of the bits, and the adversary chooses a suitable distortion to maximize their respective objective functions.

Let q_0 be the probability of a fingerprint bit being 0, with $1\text{-}q_0$ being the probability of the bit being 1. The designer chooses the value of q_0 so that the strategy space for D is $S_D = 0 \leq q_0 \leq 0.5$. The strategy space of the adversary consists of possible distortions of the multimedia that will not introduce high perceptual distortion. We model the effect of this distortion as i.i.d. noise in the fingerprint domain. Denote by p_{01} the probability of a fingerprint bit 0 changing to 1 and by p_{10} the probability of a bit with value 1 changing to 0 after the distortion. As the adversary chooses these values, his strategy space is given as $S_A = \{0 \leq p_{01}, p_{10} \leq 1\}$.

When a query multimedia is presented for identification, the designer compares its fingerprint Y with the fingerprints $X_1, X_2, \ldots, X_N$ in the database and performs a hypothesis test to determine whether it is a match or not [2]. As the main goal of the designer is to achieve a high probability of correctly identifying the multimedia, denoted as P_c, while maintaining a low probability of false alarm P_f, we use a metric that captures the tradeoff between these two quantities as a payoff function for the designer. Let $D_{KL} = KL(\, p(y|x)\, q(x) \,||\, q(y)\, q(x)\,)$ be the Kullback-Leibler (KL) divergence between the joint distribution of the bits of the query fingerprint and the reference fingerprint under the two hypotheses. As the KL divergence between the two distributions is related to the probability of making an error via the Chernoff-Stein Lemma, we set the utility function for the designer to be $U_D(q_0,p) = D_{KL}$.

For the adversary, the main goal is to prevent identification of the content while minimizing the distortion introduced into the content. The adversary needs to find an optimal tradeoff between these conflicting objectives. Thus, the utility function for the adversary consists of two terms. The first one measures the probability of evading detection, and we set this term to be equal to $-D_{KL}$. The second term penalizes the adversary for distortion introduced. We assume that the distortion in the content is manifested in the fingerprint domain as changes in the fingerprint bits, so that the overall utility function can be written as

$$U_A(q_0, p) = -D_{KL} - c_d \frac{1}{L} E[d(Y, X)],$$

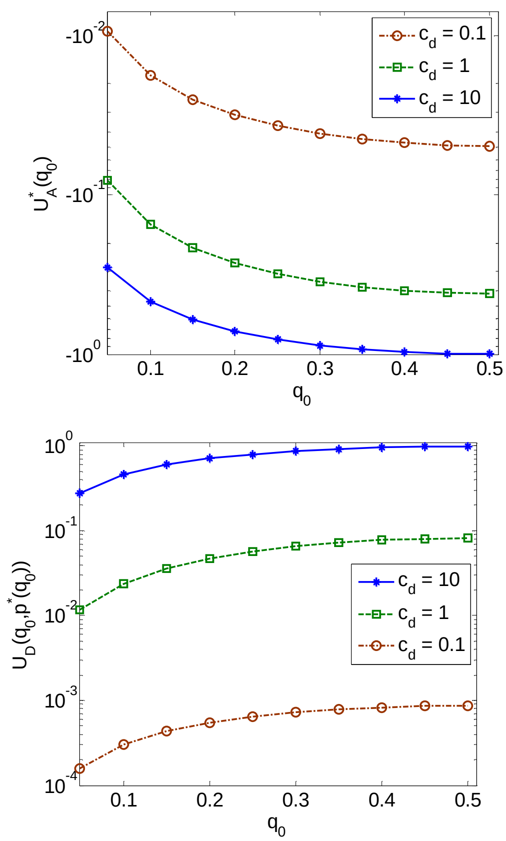
where E[$d(Y,X)$]/L is the average Hamming distance between the query and original fingerprints and c_d is a weighting factor.

3. Equilibrium Strategies

Under the above setup, the two-player game corresponds to a sequential game with perfect information [8], where the designer first chooses the fingerprint distribution and the adversary subsequently chooses the optimal distortion strategy. For each strategy of the designer, the adversary chooses the strategy that maximizes his utility function. The designer then chooses the strategy that maximizes his utility function, assuming that the adversary plays his best response strategy.

At this equilibrium point, the utilities for the adversary and the designer are shown in Fig. 1. We observe that for a given choice of q_0 by the designer, as the penalty for introducing distortion into the content increases, the payoff for the adversary reduces, as he cannot make significant changes to the fingerprint. On the other hand, the designer's payoff increases as c_d increases. This indicates that the fingerprint should be designed to make it difficult to alter the fingerprint bits without introducing high distortion into the multimedia. Furthermore, for any given c_d, the designer gets the maximum payoff when $q_0 = 0.5$, implying that the fingerprint bits should take the values 0 and 1 with equal probability.

(a) Adversary's utility function



(b) Designer's utility function

Figure 1: Equilibrium utility functions

4. Conclusions

The interaction between the fingerprint designer and the adversary seeking to evade detection in a content fingerprinting application can be modeled under the framework of game theory. In this article, we illustrated the framework using the example of binary i.i.d. fingerprints. A similar approach could be used to analyze fingerprints with correlations using the models proposed in [9]. The game-theoretic analysis provides a foundation for intuitive ideas regarding the design of binary fingerprints. This facet of the content identification problem is relatively less-explored and further study can reveal interesting insights and guidelines for designing fingerprinting algorithms.

References

[1] J. Lu, "Video fingerprinting for copy identification: From research to industry applications," in *SPIE/IS&T Media Forensics and Security*, Jan. 2009.
[2] A L. Varna, A. Swaminathan, and M. Wu, "A decision-theoretic framework for analyzing binary

hash-based content identification systems," in *ACM Workshop on Digital Rights Management*, Oct. 2008.
[3] A. L. Varna and M. Wu, "Theoretical modeling and analysis of content identification," in *IEEE Int. Conf. on Multimedia & Expo*, Jul. 2009.
[4] W.-H. Chuang, A. L. Varna, and M. Wu, "Modeling and analysis of ordinal ranking in content fingerprinting," in *IEEE International Workshop on Information Forensics and Security*, Dec. 2009.
[5] P. Moulin, "Statistical modeling and analysis of content identification," in *Information Theory and Applications Workshop*, Feb. 2010.
[6] J. Haitsma, T. Kalker, and J. Oostveen, "Robust Audio Hashing for Content Identification," in *International Workshop on Content-Based Multimedia Indexing*, Brescia, Italy, Sept. 2001.
[7] B. Coskun, B. Sankur, and N. Memon, "Spatio-temporal Transform Based Video Hashing," *IEEE Transactions on Multimedia*, vol. 8, no. 6, pp. 1190–1208, Dec. 2006.
[8] M. Osborne and A. Rubinstein, *A Course in Game Theory*, MIT Press, first edition, 2001.
[9] A. L. Varna and M. Wu, "Modeling Content Fingerprints using Markov Random Fields", *IEEE Workshop on Information Forensics and Security*, Dec. 2009.

Avinash Varna is a PhD candidate with the Department of Electrical and Computer Engineering, University of Maryland. He received the B.Tech degree from the Indian Institute of Technology Madras, Chennai and the M.S. degree from the University of Maryland.

He was awarded the Distinguished Dissertation Fellowship by the University of Maryland and the Litton Fellowship for academic Excellence. His research interests include information forensics and multimedia security. For more information about his research and publications, visit http://www.ece.umd.edu/~varna.

Min Wu received the B.E. degree in electrical engineering and the B.A. degree in economics in 1996 from Tsinghua University in Beijing, China (both with the highest honors), and the Ph.D. degree in electrical engineering from Princeton University in 2001. Since 2001, she has been with University of Maryland, College Park, where she is currently an Associate Professor.

Dr. Wu leads the Media and Security Team (MAST) at University of Maryland, with main research interests on information security and forensics and multimedia signal processing. She has co-authored two books and holds eight U.S. patents on multimedia security and communications. She is a co-recipient of two Best Paper Awards from the IEEE Signal Processing Society and EURASIP, respectively. She also received a NSF CAREER award in 2002, a TR100 Young Innovator Award from the MIT Technology Review Magazine in 2004, an ONR Young Investigator Award in 2005, a Computer World "40 Under 40" IT Innovator Award in 2007, and an IEEE Mac Van Valkenburg Early Career Teaching Award in 2009. She is currently serving as Vice President - Finance of the IEEE Signal Processing Society, Vice Chair of the IEEE Technical Committee on Information Forensics and Security, and involved in several technical committees and journals. She was elected IEEE Fellow for contributions to multimedia security and forensics. [http://www.ece.umd.edu/~minwu/]

A Game-theoretic Interpretation of P2P Swarm Equilibrium for Incentive Optimization

Suman Deb Roy and Wenjun Zeng, University of Missouri, Missouri, USA
{ debs, zengw}@missouri.edu

1. Introduction

Peer to Peer (p2p) networks are a popular medium of file sharing where a group of peers that are collectively connected to a single file represent a swarm. Unfortunately, p2p has not yet been accepted as a widespread commercial platform for multimedia content distribution. This irregularity can be accounted towards the fact that p2p protocols are easily susceptible to malicious nodes and cheating strategies. Since p2p is a collaborative environment, incentives are pivotal for peers to co-exist. Incentive optimization refers to providing enough incentive to peers to be altruistic and no incentive to engage in cheating. However, manipulation of incentives by malicious nodes [4] within the swarm results in stolen bandwidth; hence is a serious security threat.

Our research focuses on one such popular p2p protocol – BitTorrent (BT) [5], and explores previously undocumented threats which exploit the core of BT's algorithm. We identify two such exploitations and detect their cause as design flaw in the original protocol. We then develop a new protocol called *prTorrent* [1], which while similar to BitTorrent, uses an advanced unchoking algorithm (the core algorithm for peer selection in BitTorrent) that alleviates these attacks.

We also show that the reason *prTorrent* thwarts the attacks is because it captures the swarm dynamics better than BitTorrent and establishes a Nash Equilibrium (NE) within the swarm. Peers in NE abstain from engaging in cheating, since this would yield lower payoffs for them (which is the fundamental definition of a system in NE – changing the strategy would lead to lower payoffs). Another goal was to find the point at which fair peers turn into cheaters, or when the equilibrium is broken. This transformation is usually caused when there is better incentive for cheating in the swarm. Our research in [2] contains detailed analysis on conditions for maintaining the equilibrium. Finally, the design of the *prTorrent* simulator, which is easily extensible to an emulator, is described in [3].

2. Hunting for Security Holes

We discovered two vulnerabilities in the current BT file sharing algorithm that make it unsuitable for large scale content distribution and eventual adoption as a commercial platform for the same. Both these vulnerabilities exploit the BT unchoking algorithm in a game-theoretic manner to gain undue advantage. Therefore, security solutions that we develop are also game-theoretic in nature. Let us briefly understand the cause and effect of these attacks.

BitTorrent's unchoking algorithm is tit-for-tat [5] based. As a result, every move by one peer is reciprocated by another for cooperation and piece exchange. The importance of one peer to another is judged by the rarity of pieces possessed by either of them. This rarity is a discount parameter (DP), in the sense that it judges the future utility of a peer. If the DP is low, peer X can snub peer Y after it receives a piece from Y, thus thwarting Y a download opportunity. This means X does not provide Y with a chance to download since it does not wish to interact with Y again. Results show that strategic use of DP attack can provide the attackers with a 11% improvement in download speed at the cost of fair peers which suffer a reduction of 23% in download speed.

Previous research [6] has also shown that BT peers have incentive to strategically under-report pieces they possess. We found that if peers collude in under-reporting specific portions of the file, they could lead the entire swarm towards premature starvation. Premature Starvation occurs at around $0.75t$ (for a total file download time of t) whereas signs of any generic starvation [7] would not be normally visible before $0.95t$. Again, the colluding peers steal bandwidth from fair peers by under-reporting specific pieces of the file.

3. The Collaborative Solution: *prTorrent*

prTorrent is based on a simple idea: that Piece Rarity (PR) must be used when unchoking. Currently BT unchokes based on bandwidth alone. We need unchoking to be a function of both bandwidth and PR.

This would prevent PR from being a DP and thwart the first attack. For this, we define PR such that it represents the dynamics of the swarm aptly. PR is comprised of various factors from several layers of a p2p network. We call these layers – levels. Table 1 lists the PR parameters. Local availability is the availability in regards to the piece within the swarm under consideration. Global availability refers to the availability in regards to the piece within all the swarms on Internet. Completion Factor reflects the percentage of entire file downloaded by a peer. Contention is the ratio of peers to seeds in the swarm. Availability Difference is the difference of global and local availabilities.

TABLE I. PIECE RARITY PARAMETERS (Θ)

P.No.	Piece Rarity Parameters		
	Name	***Level***	***Denoted by***
1	Global Availability	Piece	g
2	Local Availability	Piece	l
3	#Upload Slots	Peer	n_u
4	Completion Factor	Peer	CF, γ
5	Contention	Swarm	$1/\beta$
6	Availability Difference	Swarm	AD

Having obtained the parameters, we use Expectation Maximization on their mixture model with the goal to find the maximum likelihood PR (σ). This gives us the PR equation for optimum incentives. Once we have PR, we plug it in a modified form of BT's unchoking equation. The original equation (1) for unchoking in BT is:

$$_{A}\chi^{B}(t) = \frac{\chi_A \cdot {}_{B}\chi^{A}(t-1)}{\sum_k {}_{K}\chi^{A}(t-1)} \qquad (1)$$

where, $_{A}\chi^{B}(t)$ represents the amount of upload bandwidth that A allocates to B in round t. We modify this equation to make room for PR. So, the *prTorrent* unchoking equation is of the form:

$$_{A}\chi^{B}(t) = \frac{\chi_A \cdot \sum_{n=1}^{j} \left[{}_{B}\chi^{A}(t-n) + 100\,\sigma_{t-n} \right]}{\sum_k {}_{K}\chi^{A}(t-1)}$$

$$where, \quad j = \begin{cases} 2 & , N_{rp} > \eta^2 \\ 1 & , N_{rp} \leq \eta^2 \end{cases}$$

The exact probability with which A will respond with cooperation in interactions depends on A's interaction history and on B's defection rate (η). The interaction history gives A the count on the number of rare pieces it has downloaded from B (denoted as N_{rp}).

Simulation results in [1] show that with prTorrent, every DP Exploiter performs worse than a fair peer in terms of download time, and prTorrent provides incentive not to under-report.

4. Maintaining the Balance

As mentioned earlier, *prTorrent* induces a behavioral pattern in p2p that dissuades peers from changing their strategy for individual benefit. This balance is critical in forcing cooperation among selfish peers and exemplifies a NE. Although previous research mentioned existence of such equilibrium in the swarm, to our best knowledge, [2] was the first to analyze conditions that *maintain* the equilibrium. Considering peer to peer interactions as a case of Prisoners' Dilemma [8] (PD) and a repeated game, our work finds the minimum number of altruistic peers required in a swarm to hinder cheating. In other words, it detects the 'tipping point' in terms of participating (fair) vs. shirking (cheating) peers that causes the payoff for cheaters to be higher or vice versa.

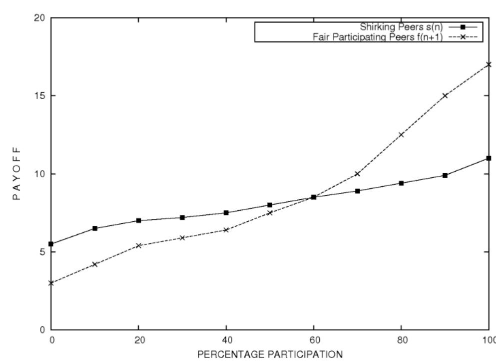


Figure 1: Nash Equilibrium for *prTorrent*

Fig. 1 shows that the overall payoff for fair peers overtakes cheating peers in prTorrent as the percentage of participation increases. In BT, the curves take opposite trend (not shown here), i.e. shirkers gain more payoff as the percentage of participation increases. This is the principal cause why in spite of existence of a NE in BT, BT still suffers from incentive manipulation. Generally, for the PD repeated game, BT achieves the worst case solution, whereas *prTorrent* achieves the collaborative solution. This collaboration alienates colluding peers and thwarts the second attack.

In other words, *prTorrent* is an Evolutionary Stable Strategy (ESS) within the swarm. An ESS is a strategy in game theory which, if adopted by a population of peers, cannot be invaded by any alternative strategy (cheating) that is initially rare.

5. Conclusion

Our work focuses on making peer-to-peer applications secure for commercial multimedia distribution. We realize that incentive manipulations are the biggest security threat in BT. To this end; we discovered two lethal vulnerabilities in BT's core unchoking algorithm. In providing solutions to these threats, we developed a new protocol called *prTorrent*, which is modeled on BT but harnesses the swarm dynamics accurately. In addition to being secure to incentive threats, *prTorrent* instill a collaborative best solution Nash Equilibrium within the swarm, which optimizes incentives.

References

[1] S. D. Roy and W. Zeng, "prTorrent: On Establishment of Piece Rarity in the BitTorrent Unchoking Algorithm," in *Inter. Conf. on Peer to Peer Computing* (P2P'09), Sept. 2009.

[2] S. D. Roy and W. Zeng, "Securing Bi-Directional Information Exchange in P2P-like Dissemination protocols through Swarm Equilibriums," submitted to *IEEE Transactions on Information Security and Forensics*.

[3] S. D. Roy, T. Knierim, S. Churchman and W. Zeng, "Design issues of the prTorrent file sharing protocol", in 7th *IEEE Conf. on Consumers Communication and Networking Conference* (CCNC'10), 2010.

[4] S. Jun and M. Ahamad, "Incentives in BitTorrent induce free riding," in *Economics for Peer-to-Peer Systems*'05.

[5] B. Cohen, "Incentives build robustness in BitTorrent," in *Proc. of Economics of Peer-to-Peer Systems*, Berkeley, California, USA, 2003.

[6] D. Levin, K. L. Curts, N. Spring and B. Bhattacharjee, "BitTorrent is an Auction: Analyzing and Improving BitTorrent's Incentives," in *ACM Special Interest Group on Data Communication*'08.

[7] F. Mathieu and J. Reynier, "Missing Piece Issue and Upload Strategies in Flashcrowds and P2P-assisted File sharing", in *Inter. Conf. on Internet and Web Applications and Services* '06.

[8] M. Feldman, K. Lai, I. Stoica and J. Chuang, "Robust incentive techniques for peer-to-peer networks," in *Proc. of the 5th ACM Conf. on Electronic Commerce,* 2004.

Suman Deb Roy is a 3rd year PhD fellow in the Center for Cyber Security Research at the University of Missouri at Columbia. His work focuses on Security including sensor security, web 2.0 security, video encryption and peer-to-peer algorithms. He has also worked in LBS and Vision projects using Android smartphones. His current research centers around semantic search. Suman has previously served as a Multimedia Encryption Engineer at Huawei Technologies, NJ. He is currently a graduate instructor at the Dept. of Computer Science.

Wenjun (Kevin) Zeng (S'94-M'97-SM'03) is an Associate Professor with the Computer Science Department of University of Missouri, Columbia, MO. He received his B.E., M.S., and Ph.D. degrees from Tsinghua University, the University of Notre Dame, and Princeton University, respectively, all in electrical engineering. His research interest is in the general areas of multimedia processing, communications, and networking.

Prior to joining Univ. of Missouri in 2003, he had worked for PacketVideo Corp., Sharp Labs of America, Bell Labs, and Matsushita Info. Tech. Lab of Panasonic Technology. From 1998 to 2002, He was an active contributor to the MPEG4 Intellectual Property Management & Protection standard and the JPEG 2000 image coding standard, where four of his proposals were adopted. He is an Associate Editor of the *IEEE Trans. on Info. Forensics & Security*, *IEEE Trans. on Circuits & Systems for Video Technology,* and *IEEE Multimedia Magazine,* and is on the Steering Committee of *IEEE Trans. on Multimedia*, of which he also served as an Associate Editor from 2005 to 2008. He is serving as the Steering Committee Chair of *IEEE Inter. Conf. Multimedia and Expo (ICME),* and has served as the TPC Vice Chair of the ICME 2009, the TPC Chair of the 2007 *IEEE Consumer Communications and Networking Conference*, the TPC Co-Chair of the *Multimedia Comm. and Home Networking Sym.* of the 2005 *IEEE Inter. Conf. Communications*. He was a Guest Editor of *the Proceedings of the IEEE's Special Issue on Recent Advances in Distributed Multimedia Communications* published in January 2008 and the Lead Guest Editor of *IEEE Trans. on Multimedia's Special Issue on Streaming Media* published in April 2004.

TECHNOLOGY ADVANCES

Mobile Video Communications

Guest Editor: Zhifeng Chen, Interdigital Communications LLC, USA
Zhifeng.Chen@interdigital.com

Global mobile data bandwidth usage increased significantly within these years. According to the recent published Cisco Visual Networking Index, mobile video traffic will exceed 50 percent, of total mobile data traffic, for the first time in 2011. Thanks to the successful convergence of more efficient video technology, e.g. H.264, improvement of air interface, e.g. LTE, smart device technology, e.g. smart phone, and prevalence of social network, e.g. Facebook, mobile video is making its biggest strides into ubiquity.

However, there are still lots of challenges to be solved in mobile video such as power consumption, delay constraint, low bandwidth and error-prone channels. This special issue aims to bring more insight from different aspects of mobile video challenges. Six papers are included which range from modeling, device design to mobile video application.

The first paper "Power Consumption Model for Scalable Video Decoding on Mobile Platform" by Zhan Ma and Yao Wang address the problems in wireless video streaming and playback for different users over different access network and with limited battery power supply in mobile handhelds. The authors introduce the complexity model for joint temporal and amplitude scalable video decoding, and then extend the proposed complexity model to power consumption model on popular ARM platform.

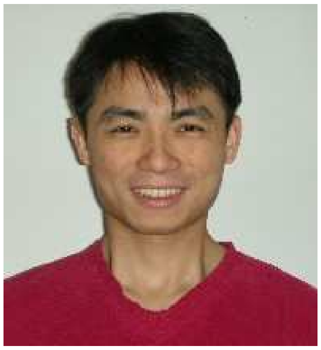
The second paper "Delay Constrained Video Transmission over Wireless Channels" by Qian Chen and Dapeng Wu review existing literature on the analysis of delay in wireless video transmission. In order to arrange the optimal time among video encoding and the rest of the delay to maximize the system performance, the authors propose a delay-rate-distortion model for video coding, which relates source distortion with encoding time.

In the third paper "Improving Energy Efficiency of DRAM Image Data Access in Video Processing" by Yiran Li and Tong Zhang, authors try to reduce the power consumption for mobile video from the perspect of chip design, i.e., reducing DRAM image data access energy consumption. The proposed 3D DRAM data storage organization can seamlessly support various motion estimation algorithms with variable block sizes. Simulation results show that the power reduction on proposed DRAM access contributes substantially to the total power reduction of the entire decoding system up to 30.4%.

The fourth paper "MobileASL: Real-Time Video Conferencing of American Sign Language over Cell Phones" by Rahul Vanam and Jaehong Chon introduce a novel low bit rate mobile video application, a real-time video conferencing system for cell phones for Deaf people who use American Sign Language (ASL). MobileASL has been ported to the HTC TyTN-II phone with consideration of different techniques: encoder speed optimization, Region-of-interest-based compression, and Battery power optimization. The field study showed that the participants preferred using MobileASL over text message for communication because they were able to see each other's expressions and reactions.

The fifth paper, titled "HTTP Live Video Streaming over Mobile Networks" by Yago Sánchez, Cornelius Hellge, Thomas Wirth, and Thomas Schierl, investigates the challenges and advances in Dynamic Adaptive Streaming over HTTP (DASH). Authors discuss the stringent delay constraint in Live DASH and also error-prone characteristics in Mobile Networks for Live DASH. Specifically, the challenges in HTTP/TCP live streaming services over LTE are investigated. Authors also discuss how the SVC technique can improve the transmission of live content with stringent delay constraints in a live streaming scenario.

In the sixth paper "The Slice Group-Based SVC Rate Adaptation Using Channel Prediction Model" by Eun-Seok Ryu and Sung Won Han, authors propose the slice group (SG)-based SVC rate adaptation using a channel prediction model for the Wi-Fi network. This method can be applied to the STB, which supports in-home wireless video streaming to multiple clients.



Zhifeng Chen received the B.E. degree from the East China University of Science and Technology, Shanghai, China, in 2001, and the M.S. and Ph.D. degrees from the University of Florida, Gainesville, Florida, in 2008 and 2010 respectively. From 2002 to 2003, he was an engineer in EPSON (China), and from 2003 to 2006, he was a senior engineer in Philips (China), both working in mobile phone system solution design. From May 2009 to Aug 2009, he was an intern in Dolby, Burbank, CA, where he had worked in error resilient rate distortion optimization. He joined the CTO office at Interdigital, San Diego, CA, in 2010, where he is currently a staff engineer working in the video coding research. His research interests include low-complexity video and image compression, perceptual video coding, Error-resilient video coding, rate-distortion optimization, rate control, cross-layer design, information theory, statistics, signal processing. He is the author of several journal and conference papers and has been awarded six patents.

Power Consumption Model for Scalable Video Decoding on Mobile Platform

Zhan Ma, Yao Wang, VideoLab, Polytechnic Institute of NYU, Brooklyn, NY 11201, USA
Zhan Ma, Samsung Telecommunications America, Richardson, TX 75082, USA
zhan.ma@sta.samsung.com, yao@poly.edu

1. Introduction

Wireless video streaming and playback (WVSP) is one of the most popular applications on mobile devices, such as Youtube on iPhone. There are two fundamental problems for a successful WVSP application. One is how to deliver the same high quality content to different users over different access network without introducing much operation overhead. The other is how to deal with the limited battery power supply for current mobile handhelds without losing much video playback quality.

We propose to use the scalable video (i.e., scalable extension of the H.264/AVC (SVC) [1]) to satisfy the diversities introduced by the underlying access networks and subscribed mobile receivers. In practice, a single scalable video stream can be easily truncated into sub-streams with different reconstruction quality levels (e.g., in terms of spatial, temporal and amplitude (SNR) resolutions (STAR)) to meet the underlying network and end-user differences. Compared with the video transcoding approach, which usually requires the powerful server to do computational intensive video transcoding, scalable video just needs a lightweight adaptor to do the bit stream adaptation. Compared with the simulcast, scalable video can reduce the total network bandwidth requirement dramatically, especially when supporting many subscribers with different network bandwidth.

This article addresses the development of power consumption model for scalable video decoding considering the joint effects of the temporal and amplitude scalability. Moreover, power consumption can be expressed as a function of the required video decoding complexity (in terms of the cycles per second or Hz) [2], i.e.,

$$P = \Phi(C), \tag{1}$$

where P and C describe the power consumption and computational complexity for video decoding respectively, and $\Phi()$ abstracts the relationship between power consumption and complexity which should be fixed for a typical hardware architecture [2]. Therefore, it is necessary to have an accurate complexity model for scalable video decoding which can be used to derive the power consumption through (1).

In the reminder of this article, we first introduce the complexity model for joint temporal and amplitude scalable video decoding, and then extend the proposed complexity model to power consumption model on popular ARM platform. We also discuss the future applications of our proposed model.

2. Complexity model for scalable video decoding

Instead of using inefficient SVC reference software – JSVM [3], we have developed our high modularized macroblock (MB) based SVC decoder targeting for the mobile handhelds. To model the complexity for SVC bit stream decoding, we implement the complexity profiler to collect the frame decoding complexity. The number of computational cycles spent in complexity profiling is less than 0.001% of the cycle number desired by the regular decoding module according to our measurement data. Hence it is negligible.

To explore the impact of temporal and amplitude scalability on the decoding complexity, we propose to apply an impact separation methodology. We normalize the raw data points (i.e., cycles per second) for any given combination of temporal and amplitude resolution, by the data points at maximum temporal and amplitude resolution (i.e., maximum complexity) respectively, and then investigate how to use appropriate analytical functional forms to model the separated normalized effects. As a result, the decoding complexity model is expressed as the product of a function of temporal resolution (i.e., frame rate) and a function of amplitude level (i.e., quantization stepsize), with three model parameters in total.

Towards this goal, the complexity model $C(q,t)$ can be written as

$$C(q, t) = C_{max} C_t(t; q_{min}) C_q(q; t), \tag{2}$$

where C_{max} is the maximum complexity demanded by decoding bit streams coded at maximum frame rate and minimum quantization stepsize, i.e., $C_{max} = C(q_{min}, t_{max})$; $C_t(t; q_{min}) = C(t, q_{min})/C(t_{max}, q_{min})$ is the normalized complexity vs. temporal resolution (NCT) under the minimum quantization stepsize q_{min}, and $C_q(q;t) = C(t,q)/C(t,q_{min})$ is the normalized complexity vs. quantization (NCQ) at any given frame rate t. Note that NCQ function $C_q(q; t)$ tells how does complexity decrease as the quantization stepsize increases beyond q_{min}, given any frame rate t; while NCT function $C_t(t;\, q_{min})$ characterizes how does complexity reduce as the frame rate decreases from t_{max}, under the q_{min}. As will shown later by experimental data, we have found that NCQ function $C_q(q;\, t)$ can be modeled by a function of q with the model parameter as a function of the frame rate t, denoted as $C_q(q;\, s(t))$ with $s(t)$ as the frame rate dependent parameter, and NCT function $C_t(t;q)$ is independent of the q, denoted as $C_t(t)$.

To see how complexity relates the temporal and amplitude resolutions, we decode the bit streams created using videos with different content activities (e.g., texture, motion, contrast, etc) and different resolutions (i.e., CIF, WVGA and 720p). Because of the space limitation, we only present the some results for CIF videos here. More details can be found in [4].

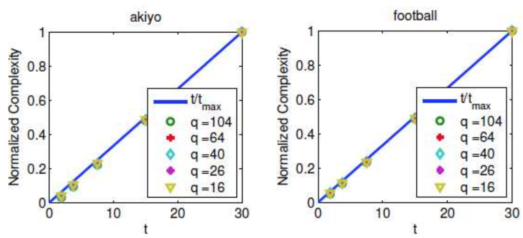


Figure 1: Illustration of NCT for different *q*. Points are profiled complexity; curves are predicted complexity using (3).

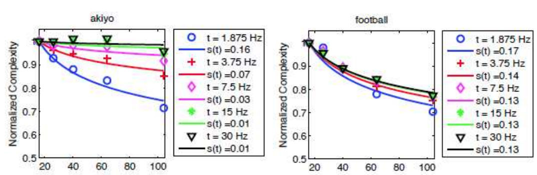


Figure 2: Illustration of NCQ for different *t*. Points are profiled complexity, curves are predicted complexity using (4).

The test bit streams are generated with five temporal (i.e., frame rates at 1.875, 3.75, 7.5, 15, 30 Hz) and five amplitude layers (i.e., QPs at 44, 40, 36, 32, 28), therefore, we can have 25 extracted bit streams. Each extracted sub stream is decoded to collect the complexity. As shown in Fig. 1, the NCT points overlap for different q and can be captured by a single curve quite well. Similarly, the NCQ curves for different t are plotted in Fig. 2. Unlike NCT, NCQ curves are frame rate dependent. However, we have found that the NCQ curves can be also predicted accurately by a power function of q with a frame rate dependent parameter $s(t)$. Thus, the overall complexity modeling work is divided into two parts, one is to develop an appropriate functional form for $C_t(t)$, so that it can model the measured NCT points accurately, the other is to devise a proper function for $C_q\,(q;\, s(t))$ to model the measured NCQ points accurately for every frame rate t. The derivation of respective $C_t(t)$ and $C_q(q;s(t))$ are explained as follows.

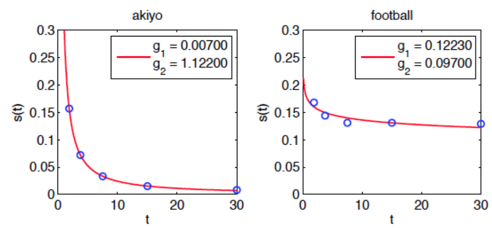


Figure 3: Power function approximation for $s(t)$, g_1 and g_2 are content dependent parameters.

2.1 Model for Normalized Complexity versus Temporal Resolution $C_t(t)$

As suggested earlier, $C_t(t)$ is used to describe the reduction of the normalized complexity as the frame rate reduces. Based on the measurement data in Fig. 1, we choose a linear function to model the NCT curves, i.e.,

$$C_t(t) = t/t_{max}, \qquad (3)$$

where t_{max} is a known constant (e.g., t_{max} = 30 Hz in our work). Fig. 1 shows the model curve using (3) along with the measured data. It can be seen that the model fits the measured data points very well.

2.2 Model for Normalized Complexity versus Quantization $C_q(q;\, s(t))$

The same to the $C_t(t)$ function, $C_q(q;s(t))$ is applied to describe the reduction of the normalized complexity as the quantization stepsize increases (i.e., corresponding to decoding less amplitude layers) at any given frame rate t. Based on the measured data in Fig. 2, we choose an inverse power function, i.e.,

$$C_q(q;\, s(t)) = (q/q_{min})^{-s(t)},\ s(t) = g_1(t/t_{max})^{-g2}, \qquad (4)$$

where $s(t)$ is the frame rate dependent parameter. Fig. 2 shows the model curve with frame rate sensitive parameter (4) along with the measured data. In addition, $s(t)$ can be well fitted by choosing proper g_1 and g_2 for different videos as shown in Fig. 3.

2.3 The Overall Complexity Model

Combing Eq. (2), (3) and (4), we can obtain the overall complexity as

$$C(q,t) = C_{\max}(q/q_{\min})^{-s(t)}(t/t_{\max}), \qquad (5)$$

where $s(t) = g_1(t/t_{\max})^{-g2}$ with g_1 and g_2 as content dependent parameters for overall complexity model. As shown in Fig. 4, we note that the model predictions fit very well with the experimental complexity points. The model parameters, g_1 and g_2 are obtained by minimizing the root mean square error (RMSE) between actual measurements and model predictions. Table I lists the parameter values, fitting errors in terms of the relative RMSE (i.e., RRMSE = RMSE/$C_{\max}$) and Pearson Correlation (PC). We see that the model is very accurate for all videos with small RRMSE and high PC.

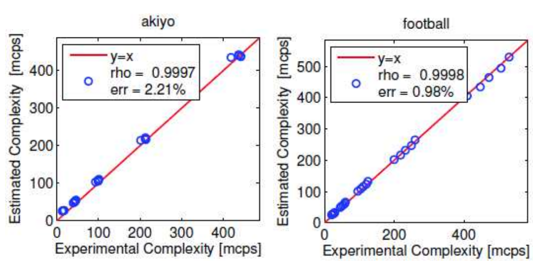


Figure 4: Illustration of complexity prediction using (5), rho and err stand for PC and RRMSE.

Table I Parameters and Model Accuracy ($C_{\max}$ is measured using MHz)

seq.	g_1	g_2	$C_{\max}$	RMSE/$C_{\max}$	PC
akiyo	0.007	1.122	443	2.21%	0.9997
city	0.056	0.571	487	1.19%	0.9997
crew	0.134	0.027	512	0.68%	0.9993
football	0.122	0.097	530	0.98%	0.9997
foreman	0.066	0.396	478	1.08%	0.9997
harbour	0.221	0.027	600	0.94%	0.9984
ice	0.067	0.263	471	1.56%	0.9996
waterfall	0.048	0.636	453	1.48%	0.9994
ave.				2.15%	0.9994

3. Power Consumption Model on ARM Platform

Popular chips, such as ARM, can support dynamic voltage frequency scaling (DVFS) according to the processor's instant workload and temperature, etc, or by user defined manner, so as to save energy. Typically, for a DVFS-capable processor, there are four kinds of power consumption: dynamic power, static leakage power, short circuit power as well as constant power. For the ARM cortex A8, which is widely used in SmartPhones, we can approximate its total power consumption [4] using

$$P(f) = k_1 f^{k2} + k_3, \qquad (6)$$

where f is the clock rate or CPU frequency in Hz (which is the same as the decoding complexity requirement), k_1, k_2, and k_3 are constant parameters for a typical platform. For our ARM cortex A8, $k_1 = 3.06\text{x}10^{-10}$, $k_2 = 3.19$ and $k_3 = 0.26$.

Combing Eq. (5) and (6), we can derive the power consumption model for scalable video decoding on ARM platform as

$$P(q, t) = k_1\, C(q, t)^{k2} + k_3, \qquad (7)$$

with k_1, k_2 and k_3 defined above. Please note that our decoding power consumption can be applied widely since the similar ARM chips are widely deployed in mobile handhelds.

4. Conclusions and Discussions

In this article, we develop a power consumption model for scalable video decoding on mobile platform. The power consumption model is extended from the decoding complexity model. Simulation results show that our proposed model can accurately predict the scalable video decoding complexity for any given temporal and amplitude combination, with small RRMSE and high PC.

As analyzed in [4], model parameters, i.e., $C_{\max}$, g_1 and g_2 can be well estimated by video content features. More details can be found in [4]. Furthermore, our power consumption model can be used to do power-rate-quality optimized video adaptation, together with the rate and quality models for scalable video [4].

References

[1] G. Sullivan, T. Wiegand, and H. Schwarz, Text of ITU-T Rec. H.264 | ISO/IEC 14496-10:200X/DCOR1 / Amd.3 Scalable video coding, ISO/IEC JTC1/SC29/WG11, MPEG08, N9574, Antalya, TR, January 2008.

[2] Z. He, Y. Liang, L. Chen, I. Ahmad, and D. Wu, "Power-rate-distortion analysis for wireless video communication under energy constraints," IEEE Trans. Circuit and Sys. for Video Technology, vol. 15, no. 5, pp. 645–659, May 2005.

[3] JSVM software, Joint Scalable Video Model, Joint Video Team, Doc. JVT-X203, Geneva, Switzerland, 29 June - 5 July 2007.

[4] Z. Ma, "Modeling of Power, Rate and Perceptual Quality of Scalable Video and Its Applications," Ph.D. dissertation, Polytechnic Institute of NYU, Jan. 2011.

Zhan Ma (S'06) received the B.S. and M.S. degrees in Electrical Engineering (EE) from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2004 and 2006 respectively, and the Ph.D. degree in electrical engineering from Polytechnic Institute of New York University, Brooklyn, in 2011. While pursuing the M.S. degree, he joined the national digital audio and video standardization (AVS) workgroup to participate in standardizing the video coding standard in China. He interned at Thomson Corporate Research lab, NJ, Texas Instruments, TX, and Sharp Labs of America, WA, in 2008, 2009 and 2010 respectively.

Since 2011, he has been with Dallas Technology Lab, Samsung Telecommunications America (STA), Richardson, TX, as a Senior Standards Researcher. His current research focuses on the next-generation video coding standardization (HEVC), video fingerprinting and video signal modeling. He received the 2006 Special Contribution Award from the national digital audio and video standardization workgroup, China for his contribution in standardizing the AVS Part 7, and 2010 Patent Incentive Award from Sharp .

Yao Wang (M'90-SM'98-F'04) received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1983 and 1985, respectively, and the Ph.D. degree in electrical and computer engineering from the University of California, Santa Barbara, in 1990.

Since 1990, she has been with the Electrical and Computer Engineering faculty of Polytechnic University, Brooklyn, NY (now Polytechnic Institute of New York University). Her research interests include video coding and networked video applications, medical imaging, and pattern recognition. She is the leading author of the textbook Video Processing and Communications (Prentice-Hall, 2001). Dr. Wang has served as an Associate Editor for the **IEEE Transactions on Multimedia** and the **IEEE Transactions on Circuits and Systems For Video Technology**. She received the New York City Mayor's Award for Excellence in Science and Technology in the Young Investigator Category in year 2000. She was elected Fellow of the IEEE in 2004 for contributions to video processing and communications. She is a co-winner of the IEEE Communications Society Leonard G. Abraham Prize Paper Award in the Field of Communications Systems in 2004. She received the Overseas Outstanding Young Investigator Award from the National Natural Science Foundation of China (NSFC) in 2005 and was named Yangtze River Lecture Scholar by the Ministry of Education of China in 2007.

Delay Constrained Video Transmission Over Wireless Channels

Qian Chen and Dapeng Wu, University of Florida, Gainesville, US
qiantrue@ufl.edu, wu@ece.ufl.edu

1. Introduction

With the development of wirelesscommunication and dramatically increase of mobile smart phone use, video streaming over wireless network becomes a big concern in both industry and academics. Examples of such applications are IPTV over mobile phones and sharing over social network via smart phones. To support these applications, future wireless networks are expected to provide quality of service (QoS) guarantees, including end-to-end delay, data transmission rate, and packet error probability. The QoS requirements pose great challenges as wireless fading channels may cause severe QoS violations.

For real time video applications, the system has a maximum tolerable end -to- end delay for each video frame, from the moment the frame captured at the encoder to the identical frame display at the decoder. Due to the fading and multipath interference of wireless network, the packet transmission error is inevitable. Therefore, certain error correction scheme needs to be employed to protect packet in error-prone channels. Forward Error Correction (FEC) coding adds redundancy among all packets for packet protection. Automatic Repeat Request (ARQ) will cause unbounded delay, and not appropriate for real time application. However, ARQ feedback combined with error control has been proved to be an alternative scheme to FEC [1][2]. An architecture employing hybrid ARQ with FEC is given in [3] for low delay wireless video transmission.

All research on wireless video transmission with delay constraint have demonstrated the relationship between delay and system performance, such as distortion, packet loss rate and throughput. [1] minimizes the source coding distortion with the rate constraint transformed from delay constraint. [2] formulates the problem as to minimize the expected end-to-end distortion with the delay constraint. [4] maximizes the packet retransmission time with delay constraint by adaptive packetization and scheduling scheme. [5] maximizes the end-to-end distortion with wireless network delay. [6][7] analyzes the system throughput with delay constraint, which is equivalent to minimizing packet error probability with a bounded buffer length. Analytical and simulation results show that delay has tremendous effect on system performance, and usually system performance is a non-monotonous decrease function with delay. How to balance the delay and performance by tuning system parameters is still a valid research problem.

Up to now, the analysis of delay in wireless video transmission can be categorized to three different methods. 1) The wireless video communication is modeled as a point-to-point system, and the end-to-end delay comes from the waiting time in encoder and decoder buffer, and the channel delay, including propagation delay and retransmission time (ARQ). It translates delay constraint in real time video transmission into a rate constraint, where the applicable rate constraints depend on future channel rates. Hence the delay requirement is equivalently to a rate control problem at the encoder. Examples in this category are [1][2][4]. 2) It is also a point-to-point system model. The delay mainly comes from the waiting time in the buffer at link layer of finite length. The delay will affect packet drop probability and delay bound violation probability. And packet error rate of the communication system is the sum of packet transmission error rate over fading channel, packet drop rate from buffer overflow and delay bound violation rate due to maximum packet delay constraint. [6] derives each of the rate and minimizes the packet error rate, which is equivalent to maximize the system throughput with delay constraint. 3) It models the video transmission over wireless networks, and the delay is due to packet transmission in a multi-hop network [5]. The end-to-end video distortion is the sum of source coding distortion and transmission distortion. It exploits a distortion-aware wireless video scheduling scheme and derives a bound on the asymptotic decay rate of system distortion. However, the source coding distortion model in [5] is too rough and not be able to get an accurate analysis of delay-distortion balance.

2. End-to-end Delay Model

Generally, in a wireless video communication system, the end-to-end delay is from the moment the frame is captured at the sender to the time it is displayed at the receiver. It is composed of the following: video encoding timeΔT_e, encoder buffer queuing time ΔT_{eb}, wireless channel transmission time ΔT_c, decoder buffer queuing time ΔT_{db} and decoding time ΔT_d [2][8], as shown in Fig.1. Each video frame is encoded and packetized to several packets for transmission. In real time application, the maximum end-to-end delay that each frame experiences is a constant.

$$\Delta T_e + \Delta T_{eb} + \Delta T_c + \Delta T_{db} + \Delta T_d = C \qquad (1)$$

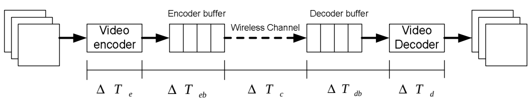


Figure 3: End-to-end Delay of a Wireless Video Communication System

All previous studies have assumed encoding time and decoding time is much less variation than the buffer and transmission delay, hence ΔT_e and ΔT_d are assumed constant. Therefore, they only focus on the system performance with buffer delay and transmission delay. However, the encoding time can be mapped to encoding complexity, and encoding complexity will affect source distortion [9]. If end-to-end delay is fixed, by tuning encoding time, the transmission distortion will also be affected as the buffer and transmission delay change. Hence, the overall system performance depends on how to assign end-to-end delay among different time components. And we are seeking the optimal time arrangement among encoding and the rest of the delay to maximize the system performance given end-to-end delay.

To achieve this, we need a model that relates encoding time and source distortion. However, there is no such model in literature to the best of our knowledge. In this letter, we propose a delay-Rate-Distortion (d-R-D) model in source encoding.

3. d-R-D in Source Coding

In [10], distortion and rate are derived as functions of residue variance σ and quantization step size Q. When Q is fixed, distortion is a function of motion estimation (ME), which can be characterized by the number of SAD, or 2D search range λ and number of reference frame θ. Hence,

$$\#SAD = \lambda \times \theta \qquad (2)$$

If we can model σ as function of λ and θ, R and D will become function of λ and θ. Meanwhile, delay d is a function of encoding complexity parameters. It has been justified in [11] ME takes the majority of the encoding complexity, and hence we use λ and θ as complexity parameters in the model. Simulation results show that residue variance σ can be approximated by exponential function to both λ and θ as follows:

$$\sigma = a\exp(b_1\lambda + b_2\theta) + c \qquad (3)$$

where a, b_1, b_2, c are all model parameters associated with each video sequence, and can be predicted from the previous frame statistics in the encoding. Fig.2 shows sample function curves between residue variance and search range λ, and Fig. 3 between residue variance and number of reference frame θ.

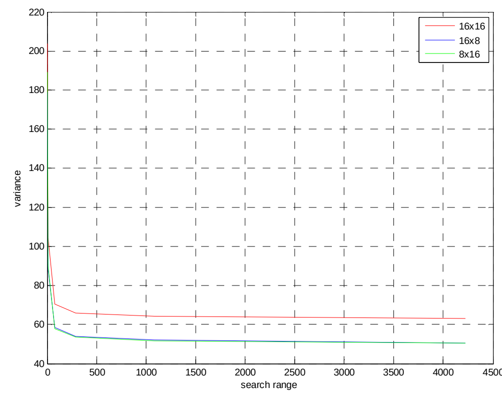


Figure 4: Residue Variance vs. Search Range of Foreman in Various Search Modes

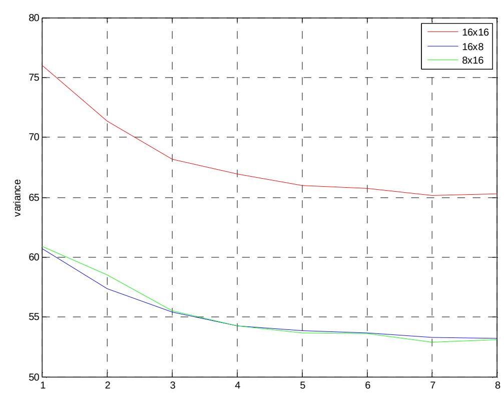


Figure 5: Residue Variance vs. Number of Reference of Foreman in Various Search Modes

Combined with source distortion and source rate model in [10][12],

$$D = \frac{\Lambda Q e^{\gamma\Lambda Q}(2+\Lambda Q-2\gamma\Lambda Q)+2-2e^{\Lambda Q}}{\Lambda^2(1-e^{\Lambda Q})} \qquad (4)$$

$$R = -P_0 log_2 P_0 + (1-P_0)\left[\frac{\Lambda Q log_2 e}{1-e^{-\Lambda Q}} - \right.$$

$$log_2(1-e^{-\Lambda Q})-\gamma\Lambda Q log_2 e+1\Big] \quad (5)$$

where

$$\Lambda=\frac{\sqrt{2}}{\sigma} \quad (6)$$

And P_0 is the probability of zero after quantization

$$P_0=1-e^{-Q\Lambda(1-\gamma)} \quad (7)$$

and γ is quantization offset. We can obtain D and R as function of λ and θ by replacing σ with (3) to (4) and (5).

Meanwhile, the encoding time of a frame can be written as:

$$\mathrm{d}=\frac{\mathrm{N}\lambda\theta\mathrm{c}_0}{\sqrt[3]{\mathrm{kP}}} \quad (8)$$

where N is the number of Macroblock (MB) in the frame, c_0 is the clock cycle of each SAD operation for certain CPU, P is CPU power, and k is the constant in CPU power model [11]. (4) means the encoding time is the total CPU cycle used in SAD operations divides the clock cycle per unit time. Fig.4 presents the relationship between source distortion (D), rate (R) and encoding delay (d) for sequence foreman. We already know from R-D theory that higher R leads to smaller D. On the other hand, it is very intuitive that when given longer encoding time, the motion estimation search range is larger. This leads to a more accurate prediction, and hence reduces the distortion.

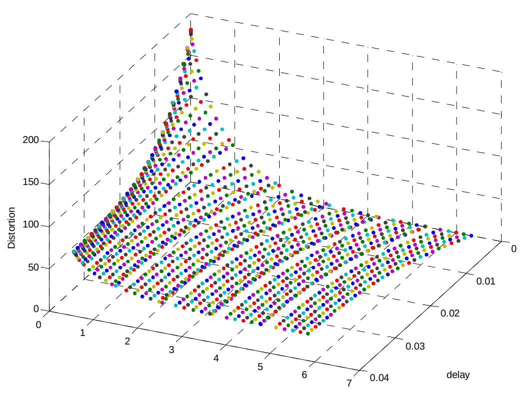


Figure 6: Source d-R-D Model for Foreman

The proposed source delay-distortion-rate model can be integrated to all previous researches on delay-sensitive wireless video communication to optimize the end-to-end system performance by arranging end-to-end delay among different components.

4. Conclusions

Video transmission over wireless channels is very sensitive to delay. We summarized the previous research methods on delay constrained wireless video communication system. They all demonstrate the balance between delay and system performance. However, these works assume encoding time is a constant and only focus on buffer delay and transmission delay. To address that encoding time has effect on overall end-to-end distortion, we propose a delay-rate-distortion model for source coding, which relates source distortion with encoding time. Hence, the optimal system performance can be obtained by arranging end-to-end delay among different components. The proposed model can be integrated into any previous research method on delay constrained wireless video transmission.

References

[1] Chi-Yuan Hsu, Antonio Ortega. A Lagrangian Optimization Approach to Rate Control for Delay-constrained Video Tranmission Over Burst-error Channels. In ICASSP'98: International Conference on Acoustics, Speech and Signal Processing, pages 2989-2992, vol.5, Seattle, WA, USA, 1998.

[2] Chi-Yuan Hsu, Antonio Ortega, and Masoud Khansari. Rate Control for Robust Video Transmission Over Burst-Error Wireless Channels. In IEEE Journal on Selected Areas in Communications, 17(5), pages 756-773, May 1999.

[3] Tu-Chih Wang, Hung-Chi Fang, and Liang-Gee Chen. Low Delay, Error Robust Wireless Video Transmission Architecture for Video Communication. In IEEE Transactions on Circuits and System for Video Technology, pages 1049-1058, 12(12), Dec. 2002.

[4] Mihaela Van Der Schaar, and Deepak S. Turaga. Cross-Layer Packetization and Retransmission Strategies for Delay-Sensitive Wireless Multimedia Transmission. In IEEE Transactions on Multimedia, pages 185-197, 9(1), Jan. 2007.

[5] Liang Zhou, Min Chen, Zhiwen Yu, Joel Rodrigues, and Han-Chieh Chao. Cross-Layer Wireless Video Adaptation: Tradeoff between Distortion and Delay. In Computer Communications, pages 1615-1622, 33(14), Sep. 2010.

[6] Xihua Dong, Xiaochen Li, and Dapeng Wu. Analysis of Packet Error Probability in Delay Constrained Communciation Over Fading Channels. In CCNC'09: IEEE Consumer Communications and Networking Conference 2009, pages 1-5, Las Vegas, NV, USA, Feb. 2009.

[7] Xihua Dong, Xiaochen Li, and Dapeng Wu. RED Theory for QoS Provisioning in Wireless Networks. Submitted to IEEE Transactions on Wireless Communications, Dec. 2009.

[8] Chi Yuan Hsu, Antonio Ortegea, and Amy R. Reibman. Joint Selection of Source and Channel Rate for VBR Video Transmission Under ATM Policing Constraints. In IEEE Journal on Selected Areas in Communications, pages 1016-1028, 15(6), Aug. 1997.

[9] Deepak S. Turaga, Mihaela Van Der Schaar, and Beatrice Pesquet-Popescu. Complexity Scalable Motion Compensated Wavelet Video Encoding. In IEEE Transactions on Circuits and System for Video Technology, pages 982-993, 15(8), Aug. 2005.
[10] Xiang Li, Norbert Oertel, Andreas Hutter and Andre Kaup. Laplace Distribution Based Lagrangian Rate Distortion Optimization for Hybrid Video Coding. In IEEE Transactions on Circuits and System for Video Technology, pages 193-205, 19(2), Feb. 2009.
[11] Zhihai He, Yongfang Liang, Lulin Chen, Ishfaq Ahmad, and Dapeng Wu. Power-Rate-Distortion Analysis for Wireless Video Communication Under Energy Constraints. In IEEE Transactions on Circuits and System for Video Technology, pages 645-658, 15(5), May 2005.
[12] Zhifeng Chen and Dapeng Wu. Rate-Distortion Optimized Cross-layer Rate Control in Wireless Video Communication. Revised and submitted to IEEE Transactions on Circuits and System for Video Technology, 2010.

Qian Chen received her M.S from Shanghai Jiao Tong University, Shanghai, China in 2008. Currently she is a PhD Candidate in Electrical and Computer Engieneering with University of Florida. From May to August 2010, she works as an intern in Technicolor, Princeton, NJ, where she worked in video coding for next generation video compression.Her research interest includes video coding and transmission, multimedia security, image and video processing. She has published several journal and conference paper and filed 4 patents. She is also the reviewer for IEEE Transaction on Circuits and System for Video Technology, Journal of Visual Communication & Image Representation, Journal of Signal Processing etc. Journal of Signal Processing etc.

Dapeng Wu received B.E. in Electrical Engineering from Huazhong University of Science and Technology, Wuhan, China, in 1990, M.E. in Electrical Engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 1997, and Ph.D. in Electrical and Computer Engineering from Carnegie Mellon University, Pittsburgh, PA, in 2003.

Since 2003, he has been on the faculty of Electrical and Computer Engineering Department at University of Florida, Gainesville, FL, where he is currently Associate Professor. His research interests are in the areas of networking, communications, signal processing, computer vision, and machine learning. He received University of Florida Research Foundation Professorship Award in 2009, AFOSR Young Investigator Program (YIP) Award in 2009, ONR Young Investigator Program (YIP) Award in 2008, NSF CAREER award in 2007, the IEEE Circuits and Systems for Video Technology (CSVT) Transactions Best Paper Award for Year 2001, and the Best Paper Award in International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine) 2006.

Currently, he serves as an Associate Editor for IEEE Transactions on Circuits and Systems for Video Technology, Journal of Visual Communication and Image Representation, and International Journal of Ad Hoc and Ubiquitous Computing. He was the founding Editor-in-Chief of Journal of Advances in Multimedia between 2006 and 2008, and an Associate Editor for IEEE Transactions on Wireless Communications and IEEE Transactions on Vehicular Technology between 2004 and 2007. He is also a guest-editor for IEEE Journal on Selected Areas in Communications (JSAC), Special Issue on Cross-layer Optimized Wireless Multimedia Communications. He will serve as Technical Program Committee (TPC) Chair for IEEE INFOCOM 2012, and has served as TPC Chair for IEEE International Conference on Communications (ICC 2008), Signal Processing for Communications Symposium, and as a member of executive committee and/or technical program committee of over 50 conferences. He has served as Chair for the Award Committee, and Chair of Mobile and wireless multimedia Interest Group (MobIG), Technical Committee on Multimedia Communications, IEEE Communications Society.

Improving Energy Efficiency of DRAM Image Data Access in Video Processing

Yiran Li and Tong Zhang, Rensselaer Polytechnic Institute, Troy, New York, USA
liy16@rpi.edu, tong.zhang@ieee.org

1. Introduction

As video processing becomes increasingly indispensible in mobile devices, its energy-efficient implementation is of great practical interest. With extremely high computational and memory requirements, video processing bridges the gap between large size of video data and limited bandwidth of most communication systems. In video processing system, the bandwidth towards the off-chip memory represents a major performance bottleneck. The increasing picture resolution requires an increase of memory bandwidth, as more pixels are fetched from off-chip memory. In addition, to achieve the high-quality, modern algorithms such as multiple reference frames (MRF) and advanced interpolation schemes are introduced. Therefore, the needed bandwidth can easily reaches several gigabytes per second, especially when the algorithms are applied to HDTV resolutions.

In motion estimation, the most resource demanding operation in video encoding, the current frame and reference frames are typically stored in large-capacity off-chip commodity DRAM, and logic chips that implement video coding typically contain on-chip SRAM as buffers to streamline the operations and reduce the off-chip DRAM access frequency. As the image resolution and frame rate continue to increase and multi-frame based motion estimation is being widely used, image data access tends to account for an increasingly significant percentage of overall video coding energy consumption. Because of significant recent developments of three-dimensional (3D) integration technology with massive vertical inter-connect bandwidth towards high manufacturability [1] 3D logic-memory integrated processing platform, which consists of one logic die and one or more high-density memory (such as DRAM) die(s), appears to be very attractive for memory-hungry applications. In general, 3D integration refers to a variety of technologies which provide electrical connectivity, e.g., with through-strata-via (TSV) [2], between multiple active device planes.

On the other hand, DRAM image data access may not benefit as much as the video processing logic operation even though the continuous technology scaling helps to reduce energy consumption. This is because, as mobile devices need to support increasingly diverse and more sophisticated functions, the storage capacity of DRAM has to accordingly increase. As a result, the DRAM die size may not reduce or even increase in spite of technology scaling down. As we will elaborate later, DRAM access energy consumption is largely dominated by the routing interconnect energy consumption that is proportional to the DRAM die size. Therefore, it is reasonable to expect that the DRAM access energy consumption will play a more and more important role in determining the overall video processing energy consumption. Taking video decoding as an example, recent work has shown that, for 1080p@30fps (frames per second) video sequences, an H.264 decoder at 130nm node only consumes 134mW, while the corresponding DRAM access power consumption is 334mW [3] Thus reducing DRAM image data access energy consumption can substantially reduce the total energy consumption of video processing systems.

2. 3D DRAM Integration

Enabled by 3D integration technologies, it is possible to improve the performance of video coding systems, in particular motion estimation, by leveraging a 3D logic-DRAM integrated heterogeneous multi-core systems as illustrated in Fig.1. It contains a heterogeneous multi-core logic die and multiple DRAM dies that provide large data storage capacity and very high data access bandwidth for the heterogeneous multi-core systems. Those common heavy duty functions can be realized as application-specific accelerators to improve the overall system energy efficiency. In current design practice, the logic chip that implements video coding fetches image frames from an off-chip commodity DRAM through high-speed chip-to-chip data links such as DDR3. As image resolution keeps increasing, such chip-to-chip data links must provide a higher bandwidth and hence tend to consume more energy. In the most straightforward manner, the use of 3D stacked DRAM can largely reduce the load on off-chip data links and hence reduce the associated energy consumption, without demanding any changes on the design of both the video codec and DRAM.

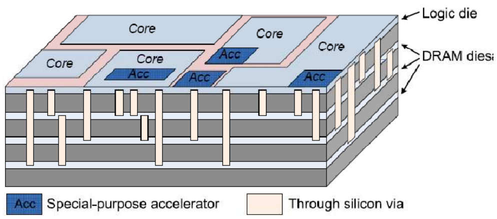


Figure 1: Illustration of 3D logic-DRAM integrated heterogeneous multi-core computing system

It is feasible and certainly preferable to store both current and a few corresponding reference frames in 3D stacked DRAM. Due to the access latency and energy consumption overhead induced by DRAM row activation, it is always favorable to make as many consecutive reads/writes as possible on the same row before switching to another row. Therefore a macroblock (MB)-by-MB mapping geared to motion estimations should be used [4], i.e., each row stores the luminance intensity data of one or more MBs.

Based upon 3D DRAM parameters in our previous work [5], we estimate the overall power consumption of memory sub-system in motion estimation realized by envisioned heterogeneous multi-core system as shown in Fig.1. And we obtain the comparison with conventional design practice. Regardless to whether 3D DRAM stacking is used or not, we assume the image frames are originally stored in an off-chip 4Gb commodity DRAM, which also stores data and instructions for other cores in the envisioned heterogeneous multi-core systems. In the case of conventional design practice, the motion estimation accelerator contains an on-chip SRAM buffer to store the current MB and a 80x80 search region in each reference frame (i.e., a total 250Kb when using 5 reference frames). We still keep the on-chip SRAM buffer in order to reduce the frequency of 3D stacked DRAM access. The estimation and comparison results are list in Table I, in which three step search algorithm is used. The access power listed in the table is the power consumed for memory read/write under the real-time constraint (i.e., 30 frames per second for HDTV1080p), including address decoder, sense amplifiers, repeaters, routing, etc. It clearly shows the energy efficiency advantages when 3D stacked DRAM is being used, mainly because the use of 3D stacked DRAM can largely reduce the frequency of off-chip 4Gb commodity DRAM access and data access to the small 3D stacked DRAM dedicated for motion estimation is much more energy-efficient than access to off-chip 4Gb commodity DRAM.

Table I. Comparison between current design practice and the design using 3D stacked DRAM with three step search

		Current practice	Design w/ 3D DRAM and on-chip SRAM
Off-chip DRAM	Access power (mW)	274.56	21.90
	I/O power (mW)	124.02	9.89
3D DRAM	Capacity (Mb)	N/A	100
	Access power (mW)	N/A	18.50
	Leakage power (mW)	N/A	31.03
	I/O power (mW)	N/A	2.96
On-chip SRAM	Capacity (Kb)	250	250
	Footprint (mm2)	0.91	0.91
	Access power (mW)	29.97	29.97
Total memory access power (mW)		428.55	114.24

3. Data Manipulation to Reduce DRAM Access Energy

DRAM is usually organized in a two-dimensional array, which consists of a number of sub-arrays. Each sub-array is a 2D matrix of memory cells and associated peripheral circuitry. On-chip H-tree distribution networks are used to route address and data throughout the DRAM chip. Using the popular memory modeling tool CACTI [6], we estimate the average single memory access energy consumption and its breakdown among different components for a 512Mb and 2Gb DRAM chip at 45nm node, as shown in Fig.2. If MB-by-MB mapping as mentioned above is used, image access mainly incurs page-mode operation in DRAM. As a result, row activation and hence bitline energy consumption will dramatically reduce and on-chip H-tree data routing energy consumption becomes completely dominant, i.e., 93% and 97% of total access energy, for 512Mbits and 2Gbits DRAM respectively.

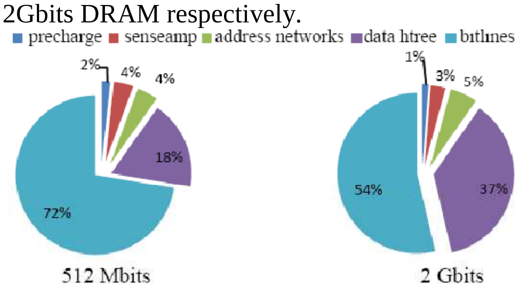


Figure 2. Estimated breakdown of single-access energy consumption for a 512Mb and 2Gb DRAM at 45nm node using CACTI

Therefore, for our interested image access in video processing, it is of paramount importance to reduce the energy consumed by data routing through H-tree in DRAM. The dynamic energy consumption of interconnect wires can be expressed as

$$E = (\alpha_s \cdot (C_s + C_L) + \alpha_c \cdot C_c) \cdot V_{dd}^2,$$

where C_s, C_L and C_c are wire self-capacitance, load capacitance, and coupling capacitance, respectively; α_s and α_c represent the self-transition and coupling-transition activity factors; and V_{dd} is supply voltage. As DRAM technology continues to scale down, the increase of wire aspect ratio and the decrease of wire pitch result in rapid increase of coupling capacitance, and therefore coupling-induced energy consumption becomes increasingly important.

In order to explicitly exploits the rich temporal and spatial data correlation in image data, we propose Gray coding to represent the pixel value to ensure small number of bit changes when pixel values are close, therefore reduce the self-transition on data bus. Intuitively, if bits transferred on adjacent wires have a larger correlation, the associated coupling-transition activity tends to be relatively small. Hence, with the objective to increase the correlations of bits transferred on adjacent wires, we propose to interleave the pixels at the bit-level. Adjacent wires carry the same-position bits, which can increase the bit correlations between adjacent wires and hence reduce the coupling-induced energy consumption.

Fig. 3 shows the average DRAM energy consumption to read or write both luminance and chrominance data of a 16x16 macroblock under various video sequences, where different combinations of these data manipulation techniques are considered. The DRAM modeling

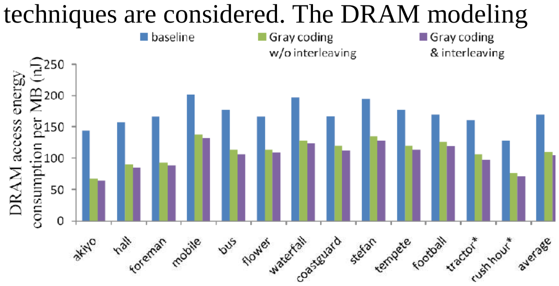


Figure 3: DRAM access energy consumption per macroblock in the case of 2Gb DRAM.

is carried out with CACTI with a capacity of 2Gbits at 45nm node. Binary coding without interleaving is the baseline for comparison. The results clearly demonstrate the effectiveness of our proposed techniques.

To further show the effectiveness of these techniques in the context of entire video decoding system, we estimate the corresponding DRAM power consumption under different video format when using either the baseline or the combination of both Gray coding and interleaving. Moreover, we extract the video decoder core power consumption based on the results presented in [7]. The decoder is implemented at 0.18μm technology node. The results are listed in Table II, in which the power is in the unit of mW. The power reduction on DRAM access contributes substantially to the total power reduction of the entire decoding system.

Table II. Power consumption (mW) comparison and saving of entire video decoding.

video format	decoder core	2Gb DRAM		power saving
		baseline	proposed	
QCIF@15fps	0.125	0.50	0.31	30.4%
D1@30fps	12.4	13.69	8.36	20.4%

4. Conclusions

Our work concerns how to reduce DRAM image data access energy consumption in video processing. With 3D memory stacking, memory access may no longer be a bottleneck for video encoding, in particular for motion estimation. We develop one specific 3D DRAM data storage organization with good energy efficiency and can seamlessly support various motion estimation algorithms with variable block sizes. We also propose to use simple yet effective data manipulation techniques to exploit the spatial and temporal correlation to reduce DRAM data routing energy consumption. Using CACTI based DRAM modeling and video processing as a test vehicle, we quantitatively demonstrated the effectiveness of these design techniques.

References

[1] J.-Q. Lu. 3-D Hyperintegration and Packaging Technologies for Micro-Nano Systems. In Proc. of the IEEE 97, pages 18-30, 2009.
[2] Z. Xu, J.-Q. Lu, "High-speed design and broadband modeling of through-strata-vias (TSVs) in 3D integration," In IEEE Trans. on Components, Packaging and Manufacturing Technology, vol. 1, no. 2, pages 154-162, Feb. 2011.
[3] D. Zhou, Z. You, J. Zhu, J. Kong, Y. Hong, X. Chen, X. He, C. Xu, H. Zhang, J.Z Hou, N. Deng, P. Liu, and S. Goto. A 1080p@60fps multistandard video decoder chip designed for power and cost efficiency in a system perspective. In IEEE Symposium on VLSI Circuits, pages262-263, 2009.
[4] E. Jaspers, P. de With. Bandwidth reduction for video processing in consumer systems. In IEEE Trans. on Consumer Electronics , vol.47, no.4, pages 885-894, 2001
[5] Y. Li, Y. Liu, and T. Zhang. Exploiting Three-Dimensional (3D) Memory Stacking to Improve Image Data Access Efficiency for Motion Estimation Accelerators. In Elsevier Journal on Signal Processing: Image Communication, vol. 25, issue 5, pages 335-344, 2010

[6] CACTI: An integrated cache and memory access time, cycle time, area, leakage, and dynamic power model, http://www.hpl.hp.com/research/cacti/.
[7] T. Liu, T. Lin, S. Wang, W. Lee, J. Yang, K. Hou, and C. Lee. A125μw, fully scalable mpeg-2 and H.264/AVC video decoder for mobile applications. In IEEE Journal of Solid-state circuits, vol. 42, no. 1, pages 161-168, Jan 2007.

Yiran Li received the B.S. degree in Electronics Engineering from Shanghai Jiaotong University, China, in 2003 and the M.S. degree in Electrical Engineering from University of Florida, Gainesville, in 2007. He is currently working toward Ph.D. degree at Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY.
From 2003 to 2006, he was with SONY Corporation, Mobile Products R&D Group, Tokyo, Japan. In 2008, he worked at Marvell Semiconductor, Inc., Santa Clara, CA. His research interests include image and video processing, memory architecture design, VLSI architecture and circuit design for multimedia processing.

Tong Zhang (M'02–SM'08) received the B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 1995 and 1998, respectively, and the Ph.D. degree in electrical engineering from the University of Minnesota, Minneapolis, in 2002. He is currently an Associate Professor with the Department of Electrical, Computer, and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY.
His research activities span over circuits and systems for various data-storage and computing applications.
He currently serves as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II and the IEEE TRANSACTIONS ON SIGNAL PROCESSING.

MobileASL: Real-Time Video Conferencing of American Sign Language over Cell Phones

Rahul Vanam and Jaehong Chon, Department of Electrical Engineering, University of Washington, Seattle, WA
{rahulv, jaehong}@u.washington.edu

1. Introduction

Video cell phones have provided the Deaf Community with the mobility to communicate with each other using Sign Language. The MobileASL team at the University of Washington and Cornell University has developed a real-time video conferencing system for cell phones for Deaf people who use American Sign Language (ASL) [1]. ASL is the native language of Deaf people in the US. This letter gives an overview of the MobileASL project.

The compression efficiency of H.264 encoder makes it suitable for use in mobile devices that operate on low bandwidth. Also, many cell phones have hardware H.264 decoders, which allow fast decoding. MobileASL uses x264 [2], an open source H.264 video encoder. Compared to the JM reference encoder, x264 is 50 times faster while providing comparable rate-distortion performance [3]. MobileASL was initially ported to the HTC TyTN-II phone, which has a front facing camera, and uses a Window Mobile 6.1 operating system, and has a 400 MHz Qualcomm MSM7200 processor. MobileASL operates on both WiFi and 3G networks. It uses Network Address Translator (NAT)-enabled protocol to allow the phones to connect to each other and to transmit the video over UDP. Figure 1 illustrates the MobileASL phone.

2. Encoder speed optimization

There are several challenges associated with our system which include high encoding complexity, maintaining good video quality, and low battery life. Since information is conveyed by both hand signs and facial gestures, a low frame rate due to a slow encoder will reduce the intelligibility of the video. The following methods were used to speedup the encoder: (1) writing optimized assembly instructions for motion estimation, mode decision, and transform operations of the encoder [4]; (2) down-sampling the raw video resolution to 96x80 before encoding [4]; and (3) tuning the encoder parameter settings both heuristically and with distortion-complexity optimization algorithms [5][6]. This improved the frame rate from 5 fps to 15 fps. Studies have shown that a minimum of 10 fps is required for maintaining ASL intelligibility [7][8], thereby indicating that MobileASL delivers intelligible videos.

Figure 1. A screen shot of the MobileASL system running on the HTC TyTN-II cell phone. The author and his colleague are shown signing to each other.

3. Region-of-interest-based compression

MobileASL can operate at bitrates as low as 30 kb/s to support video transmission over low bandwidth GPRS networks. At such low bitrates, the video suffers from blurriness and occlusions, making the video unintelligible. Therefore, a region-of-interest (ROI) encoding scheme is chosen to improve video quality [4]. This involves using a well-known low complexity skin detection algorithm [8] to identify skin macroblocks, and encoding them using a smaller quantization step size. This approach improves the PSNR by up to 2.3 dB in ROI.

4. Battery power optimization

Running video conferencing on cell phones can shorten battery life, due to encoder's complexity and power consumed by the display. To reduce the power consumed by the encoder, several schemes were developed, which involves detecting frames that are signing and listening, and dropping frames [10] and/or reducing frame resolution for listening frames [11]. Frames are classified as signing/listening by thresholding the difference of consecutive frames [10]. During listening, the signer does not convey much information, and therefore modifying these frames do not affect intelligibility. On an average, the battery life when running MobileASL without optimization is 284 minutes. Reducing the listening frame rate to 1 fps improves battery life by 23 minutes, and applying both frame dropping and downsampling improves battery life by 31 minutes. Dimming the backlight of the display during listening segments of the video together with frame dropping improves battery life by 54 minutes [12].

5. Field Study

To test the usability of MobileASL, a field study was conducted in 2010 with 11 participants from the Summer Academy for Advancing Deaf & Hard of Hearing in Computing. Figure 2 shows two participants conversing through MobileASL during the field study. Each participant was given a MobileASL phone for three weeks. The phones were programmed to popup a questionnaire after the end of few calls to ask about the person they talked to (e.g., friend, parent, etc.), where a call was made (e.g., on a bus, University, etc.), the purpose of the call, or the quality of the call. At the end of the field study the participants said that they preferred using MobileASL over text message for communication because they were able to see each other's expressions and reactions. They also mentioned that the phone used for MobileASL (HTC TyTN-II) was too heavy and its battery life for video calls was too short [13].

Figure 2. Two participants conversing through MobileASL during the field study in 2010.

6. Conclusion

MobileASL provides real-time sign language communication on cell phones over the U.S. cellular network. Since cell phone technologies such as processor speed, weight, and network speed are evolving continuously, we plan to expand MobileASL to be available on Android devices. We are also going to incorporate access to video relay service so that ASL users can communicate with non-ASL users through mobile video call.

References

[1] MobileASL, http://mobileasl.cs.washington.edu/.

[2] x264, http://www.videolan.org/developers/x264.html

[3] L. Merritt and R. Vanam. "Improved rate control and motion estimation for H.264 encoder," in *Proceedings of ICIP*, volume 5, pages 309–312, Sept. 2007.

[4] J. Chon, N. Cherniavsky, E. A. Riskin, and R. E. Ladner, "Enabling access through real-time sign language communication over cell phones," in *Proceedings of the 43rd Asilomar Conference on Signals, Systems, and Computers*, Nov. 2009.

[5] R. Vanam, E. A. Riskin, S. S. Hemami, and R. E. Ladner, "Distortion-complexity optimization of the H.264/MPEG-4 AVC encoder using the GBFOS algorithm," in *Proceedings of the IEEE Data Compression Conference*, pp. 303-312, Snowbird, Utah, March 2007.

[6] R. Vanam, E. A. Riskin, and R. E. Ladner, "H.264/MPEG-4 AVC encoder parameter selection algorithms for complexity distortion tradeoff," in *Proc. of the IEEE Data Compression Conference*, pp. 372-381, Snowbird, Utah, Mar. 2009.

[7] C. M. Reed, L. A. Delhorne, N. I. Durlach, and S. D. Fischer, "A study of the tactual and visual reception of fingerspelling," *Journal of Speech and Hearing Research*, no. 33, pp. 786-797, Dec. 1990.

[8] W. W. Woelders, H. W. Frowein, J. Nielsen, P. Questa, and G. Sandini, "New developments in low-bit rate videotelephony for people who are deaf," *Journal of Speech, Language, and Hearing Research*, pp. 1425-1433, Dec. 1997.

[9] S. L. Phung, A. Bouzerdoum, and D. Chai, "Skin Segmentation Using Color Pixel Classification: Analysis and Comparison," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 148–154, January 2005.

[10] N. Cherniavsky, A. Cavender, R. E. Ladner, and E. A. Riskin, "Variable frame rate for low power mobile sign language communication," *Assets 2007*, pp. 163-170, Tempe, AZ, Oct. 2007.

[11] J. Tran, T. Johnson, J. Kim, R. Rodriguez, S. Yin, E. A. Riskin, R. E. Ladner, and J. Wobbrock, "A Web-Based User Survey for Evaluating Power Saving Strategies for Deaf Users of MobileASL," *in proceedings of ASSETS 2010*, Orlando, FL, October 25-27, 2010.

[12] R. Vanam, Rate-Distortion-Complexity Optimization of Video Encoders with Applications to Sign Language Video Compression. Ph.D. Thesis, Electrical Engineering Department, University of Washington, Seattle, 2010.

[13] J. Kim, J.J Tran, T.W. Johnson, R. Ladner, E. Riskin, and J. Wobbrock, "Effect of MobileASL on communication among deaf users." Extended Abstracts of the ACM Conference on Human Factors in Computing Systems (CHI '11). New York: ACM Press.

Rahul Vanam received the B.E. degree in electronics and communication engineering from the Bangalore University, Bangalore, India, in 2000, and the M.S.E.E. degree from the New Mexico State University, Las Cruces, in 2005, and the Ph.D. degree in electrical engineering from the University of Washington, Seattle, in 2010. He is currently a Staff Engineer in the CTO group at InterDigital Communications, San Diego, CA, working on video compression and video processing research and development. During graduate school, he interned with Microsoft Research, Redmond, WA, Thomson Corporate Research, Princeton, NJ, and Nvidia, Inc., Santa Clara, CA. From 2000 to 2003, he was a DSP Engineer at Wipro Technologies, Bangalore, where we worked on audio codecs. His research interests include region-of-interest-based video compression, content adaptive video processing, and perceptual audio coding. He received the best student paper award at the 2009 Western New York Image Processing Workshop.

Jaehong Chon received the B.S. degree in Electrical Engineering from the Konkuk University, Seoul, Korea, in 1995, and the M.S. degree in Electronic Engineering from the Konkuk University, Seoul, Korea, in 1998. After that, he joined Samsung Electronics to develop third generation cellular systems until 2006. He is currently working toward the Ph.D. degree in Electrical Engineering, University of Washington, Seattle, WA.
His research interests include image and video processing, video compression, and video transmission over wireless networks.

HTTP Live Video Streaming over Mobile Networks

Yago Sánchez[1], Cornelius Hellge[1], Thomas Wirth[2], and Thomas Schierl[1]
Multimedia Communications[1] and Wireless Communication and Networks[2]
Fraunhofer HHI, Einsteinufer 37, 10587, Berlin, Germany
{yago.sanchez, cornelius.hellge, thomas.wirth, thomas.schierl}@hhi.fraunhofer.de

1. Introduction

In recent years multimedia delivery over the Internet has sharply increased becoming the main bandwidth consumer within the network. Parallel to this increase, significant improvements in mobile networks have lead to the apparition of high speed access networks such as 3GPP's High Speed Downlink Packet Access (HSDPA) and the emerging Long Term Evolution (LTE) networks.

With the improvements in mobile networks IP services are expected to be a ubiquitous fact of the daily life. Recent studies expect that consumption of multimedia content, especially video streaming, is going to continue increasing [1], which may also be a result of the advances in mobile networks. In fact, in [2] it has been reported that about the 50% of the data traffic in mobile networks is video data and it is expected that two-thirds of the world's mobile data traffic will be video by 2015.

HTTP streaming is one of the promising multimedia applications that has emerged in the last years and has had an incredible acceptance by the market, which is evident by the standardization activities on adaptive HTTP streaming carried out by different standardization bodies, such as MPEG [3] and 3GPP [4] or proprietary solutions such as IIS Smooth Streaming [5] and HTTP Live Streaming [6].

Although media streaming has been associated previously with RTP/UDP due to its lower latency, relying on HTTP/TCP for media delivery has shown to be a very valuable solution for scenarios where extremely stringent delay constraints are not considered, since traversal problems within NAT and Firewalls, typical with RTP/UDP, are not present.

Furthermore, existing HTTP caching infrastructure can be re-used, which relieves servers and reduces the peak traffic in the core network. A lower peak traffic allows to reduce the over provisioning of the core network, which reduces the network and delivery costs. Traversal problems may be less problematic due to the system integration of the dedicated mobile devices. However, HTTP streaming is expected to have a great success as over the top (OTT) services. Hence, it is worth to investigate how to improve HTTP streaming in mobile environments. Furthermore, both service providers and operators benefit from the reduced delivery costs over mobile networks obtained by caching, which makes HTTP streaming a very attractive service.

In this article, we briefly introduce HTTP streaming, as well as recent advances in mobile networks and we point out potential areas for improvement for a more challenging scenario: live video streaming over HTTP. We further discuss how the scalable extension of H.264/AVC [7], i.e. the Scalable Video Coding (SVC) [8], can improve the transmission of live content with stringent delay constraints in a live streaming scenario.

2. Dynamic Adaptive Streaming over HTTP (DASH)

DASH [3] is an emerging MPEG standard, which defines a format for multimedia delivery over HTTP. It basically consists of two elements: the format of the media to be downloaded and the description of the media to be downloaded. Existing proprietary solutions are based on a similar approach.

The media format is basically structured in typically small time intervals of video, called segments, which if continuously downloaded allow for a continuous representation of the media. Furthermore, usually different representations, e.g. encodings, of the media at different bitrates are available at the server allowing for a user-driven adaptation, where users select representations based on the observed network throughput. Download of segments of different representations for different time intervals is allowed resulting in a perfectly playable media, if all switching constraints presented in the Media Presentation Description (MPD), described below, are followed.

In DASH, the description of the format is given by the MPD. The MPD is an XML document, which describes the data and especially the segments available at the server. Using the MPD the clients have the necessary information to make the requests which fit their network throughput or their requirements.

In DASH the clients are responsible for performing adaptation. Based on the interests of the users, equipment capabilities and current status of the network, DASH clients have to select the representation(s) described in the MPD, which match best the necessities/capabilities of the clients. An example of DASH architecture is shown in Figure 7.

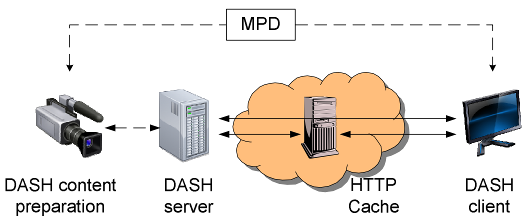


Figure 7: Example of DASH architecture

3. Considerations on Live DASH

Adaptive HTTP streaming of live content becomes a more challenging scenario than Video on demand (VoD) services, especially when very stringent delay constraints have to be fulfilled. In HTTP-based streaming services, clients typically rely on a sufficient pre-buffering to overcome the throughput variations.

Streaming of live content, usually tolerates a small delay to the time the live event occurs, which is typically less than 10 seconds. Therefore, such clients typically play the content slightly behind the live signal, allowing them to build up a sufficient safe-guard interval to overcome variations in the network. However, some scenarios, such as sport events, which are watched by large groups of the population do not tolerate such delays, since HTTP clients would receive the content significantly later than other clients do via typical broadcast channels. The users may perceive some events after becoming aware of them through other means, e.g. the exclamation of spectators in the vicinity after a soccer goal. Another possible scenario is interactive services where the spectators may have the possibility to call in real-time to the delivered show. In such scenarios a low system latency is highly desirable.

Since HTTP relies on TCP, with continuously varying throughput due to TCP's congestion control, small pre-buffer values may increase the probability of playout interruptions. In the case with small pre-buffers fast adaptation would be required to prevent these interruptions, if unexpected drastic throughput reductions happen. Therefore, there is a tradeoff between time behind the live signal and guaranteeing playout interruptions or having the requirement to support fast adaptation. In case of an as-close-to live as possible service, delays have to be kept at very low values and therefore, the adaptation has to adjust in a very fast way to variations within the network.

4. Challenges in Mobile Networks for Live DASH

In mobile networks such as in HSDPA [9] and LTE [10] TCP usage imposes additional challenges. Beside packet dropping occurring in the core network as result of competing traffic and queue overflows, packet loss may occur on the air interface. In [11] e.g. the goodput of TCP NewReno in mobile networks is analyzed, where frequent handovers may occur as a consequence of shadowing and affect the TCP throughput, since packets may be received out of order.

For LTE, different improvements have been introduced. Moving to Orthogonal Frequency-Division Multiple Access (OFDMA) in combination with Multiple-Input Multiple-Output (MIMO) enhancements and migration from circuit-switch to packet-switch networks has resulted in a mobile network that achieves peak throughputs up to 150/300 Mbps for LTE Rel. 8 with 2x2/4x4 MIMO. One of LTE's key achievements is the fulfillment of the ITU-R [16] latency requirements with a delay below 50 ms on the control plane and below 5 ms on the user plane, essential for a low end-to-end delay.

LTE implements fast retransmission mechanisms: automatic repeat requests (ARQ) and hybrid ARQ (HARQ) mechanisms at physical layer (PHY) and medium access control (MAC) layers, which requires fast re-ordering at the receiver. Thus, additional jitter and delay may be introduced by reorder buffering resulting in performance degradation for real-time TCP services, especially if HTTP/TCP video services are not identified and run over-the-top as best-effort service. TCP performance during handover in LTE is evaluated in [12] and it is shown that special packet forwarding techniques and packet reordering are necessary to achieve high TCP performance.

In addition, LTE introduces decentralized scheduling and multi-user radio resource management (RRM) at the base station, the evolved NodeB (eNB). The decentralized approach requires the design of new robust cross-layer scheduling algorithms with QoS support in order to realize end-to-end QoS for different traffic services, such as HTTP/TCP live streaming.

Furthermore, capacity will be enhanced by introducing smaller cells, e.g. Femtocells [13] with a full frequency reuse which means that neighboring base stations operate at the same carrier frequency. Inter-cell interference may lead to higher packet error rates resulting in TCP performance degradation. Therefore, more advanced RRM concepts with active interference management techniques, e.g. interference coordination, will be addressed by LTE's evolution LTE-Advanced.

Finally, with the increasing demand for HTTP/TCP live streaming services over LTE, the performance of backhaul capacity and delay become more and more important. The performance of new core network entities like the evolved packed core (EPC) and IP multimedia subsystem (IMS) and their impact on the TCP performance have to be evaluated.

5. SVC-based Adaptive HTTP Streaming

Solutions at upper layers can also be applied to overcome throughput variations. One promising solution is to use the Scalable Video Coding (SVC) [8] in combination with Adaptive HTTP Streaming. SVC allows for defining different representations (cf. section 2) within a single bitstream, by defining subsets of the complete SVC stream, which correspond to a video at a specific framerate, resolution and fidelity at a given bitrate.

Solutions relying on single layer codecs base their operation on the well-known Bitstream Switching [15] approach, where different encodings, in DASH representations, are offered at different bit rates. Decisions for switching between representations are based on previous throughput observation. However, with the challenges in mobile networks mentioned in section 4, such as multi user resource management, the TCP throughput may be drastically reduced unexpectedly. This reduction in combination with small pre-buffer values may result in buffer starvation and playout interruption due to an unexpected longer download time of the requested representation.

The challenge of unexpected throughput variations can be addressed by a prioritization of the downloaded media based on SVC, as presented in [14]. How to optimally schedule the SVC data depends on the current channel quality, delay constraints and the representations contained in the SVC stream.

When performing DASH with SVC, sub-representations (layers) are requested sequentially and downloaded. Adjustment of the bitrate of the requested representation is still possible during download due to the hierarchical download and refinement process of DASH with SVC, where requests to higher layers may be omitted when throughput reductions are noticed. Therefore, wrong estimations are less probable than with Bitstream Switching and thereby playout interruptions can be avoided more effectively.

It is also expected that SVC reduces the costs of the service providers. In [17] it has been shown that SVC enhances the cache performance and reduces the traffic in the core network in a VoD scenario. Although the reported gains are thought to be higher than in a live streaming scenario, SVC should still result in a higher cache-hit-ratio and reduced traffic in the core network, i.e. backhaul in mobile networks. Therefore, lower costs and higher benefits are expected for the service providers.

References

[1] Cisco White Paper. Cisco Visual Networking Index: Forecast and Methodology, 2009-2014, June 2010

[2] Cisco White Paper. Cisco Visual Networking Index: Global Mobile Data-Traffic Forecast Update, 2010–2015

[3] "Information technology -- MPEG systems technologies -- Part 6: Dynamic adaptive streaming over HTTP (DASH)", ISO/IEC DIS 23001-6, 2011

[4] 3GPP; Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs (Release 9); 3GPP TS 26.234 V9.3.0 (2010-06): Adaptive HTTP Streaming

[5] A. Zambelli. "IIS smooth streaming technical overview". Microsoft Corporation, 2009.

[6] HTTP Live Streaming Architecture. Technical report, Apple Inc., 2010.

[7] T. Wiegand, G. J. Sullivan, G.Bjøntegaard, and Ajay Luthra "Overview of the H.264/AVC Video Coding Standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no.7, 2003

[8] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of the H.264/AVC standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no.9, pp 1103–1120, 2007

[9] 3GPP; Physical layer; General description (Release 8); 3GPP TS 25.201 V8.1.0 (2008-05): Adaptive HTTP Streaming

[10] 3GPP; Evolved Universal Terrestrial Radio Access

(E-UTRA) and evolver Universal Terrestrial Radio Access Network (E-UTRAN); Overall description (Release 8); 3GPP TS 36.300 V8.12.0 (2010-03): Adaptive HTTP Streaming

[11] S. Sharma, D. Gillies, W. Feng, " On the Goodput of TCP NewReno in Mobile Networks", ICCCN 2010

[12] A. Tacz, A. Temesvary, N. Reider, " Handover Performance in 3GPP Long Term Evolution (LTE) Systems", Mobile and wireless Communications Summit 2007

[13] W. Mulder and T. Wirth, "Are we ready for the femtolution?" IEEE COMSOC MMTC E-letter, vol. 5, no. 5, pp. 10–13, Sep. 2010

[14] T. Schierl, Y. Sanchez, R. Globisch, C. Hellge, and T. Wiegand, "Priority-based Media Delivery using SVC with RTP and HTTP streaming", Springer Multimedia Tools and Application, September 2010

[15] Stockhammer T, Walter M, Liebl G (2005) Optimized H. 264-Based Bitstream Switching for Wireless, ICME 2005

[16] ITU-R, Report M.2134, Requirements related to technical performance for IMT-Advanced radio interface(s), Approved in Nov 2008

[17] Y. Sánchez, T. Schierl, C. Hellge, T. Wiegand, D. Hong, D. De Vleeschauwer, W. Van Leekwijck, Y. Le Louedec. "iDASH: improved Dynamic Adaptive Streaming over HTTP using Scalable Video Coding", ACM MMsys, 2011

Yago Sánchez de la Fuente received his MSc.-Ing. degree in telecom-munications engineering from Tecnun-Universidad de Navarra, Spain, in September 2009. In 2008-2009 he carried out his master-thesis on P2P application layer multicast using SVC at the Fraunhofer HHI. He is currently working as a researcher at the Technische Universität Berlin in the Image Communication Group of Prof. Thomas Wiegand and is a guest researcher at Fraunhofer HHI.

Cornelius Hellge received his Dipl.-Ing. degree in me dia technology from the Illmenau University of Technology, Germany, in July 2006. He is affiliated with the Technische Universität Berlin and at the same time he is a guest researcher at Fraunhofer HHI. He contributed to different standardization activities (e.g., DVB, 3GPP) His main research topic is the joint use of channel and source coding.

Thomas Wirth received a Dipl.-Inform. degree in computer science from the Universität Würzburg, Germany, in 2004. In 2006 he joined Fraunhofer HHI. His research interests are in the field of multi-user radio resource management for LTE-Advanced, including QoS-aware scheduling for MIMO-OFDMA systems, cross-layer design for layered encoded video transmission over LTE, LTE-Advanced relaying, SDR real-time implementations and field trials for LTE-Advanced.

Thomas Schierl is head of the Multimedia Communications Group of the Image Processing Department of Fraunhofer HHI, Berlin. His research group is working on media content delivery techniques for different distribution networks, from mobile broadcast to fixed wire-line IPTV, for a variety of media coding techniques such as SVC or 3D video. Thomas also represents Fraunhofer HHI's media transport expertise in different standardization committees such as MPEG, IETF, 3GPP and DVB.

The Slice Group-Based SVC Rate Adaptation Using Channel Prediction Model

Eun-Seok Ryu, Georgia Institute of Technology, USA,
Sung Won Han, University of Pennsylvania, USA.
esryu@gatech.edu, hansungw@mail.med.upenn.edu

1. SG-Based SVC Rate Adaptation

This paper proposes a new scalable video coding (SVC) rate adaptation method over a WIFI network studied in home gateway project with channel prediction model. To concentrate the rate adaptation itself, the error protection method using Raptor forward error correction (FEC) code is not explained in this paper though it was used in the project. The referenced paper [1] explains overall system architecture and components.

The H.264 SVC technique enables the server to adapt bandwidth of multiple clients by extracting optimal video layers from pre-encoded multiple layered video according to the client's bandwidth. It also has an error resilient feature with flexible macroblock ordering (FMO), which restructures the ordering of the representation of the fundamental regions (macroblocks) in picture. With FMO technique, the encoder classifies each slice group (SG) according to its priority. For example, the region in the center area is normally more important than the other areas. Thus, it is grouped as a high-priority slice group and prior to the other slice groups [2-3]. Besides, the proposed algorithm sets a higher priority to the slice group of the base layer because losing a base layer bitstream causes severe errors to enhancement layers in the decoding process.

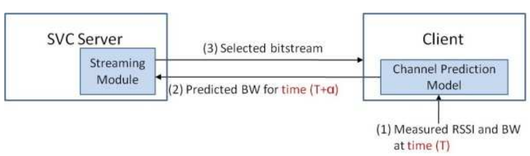


Figure 1: Conceptual procedures for the channel prediction-based rate adaptation.

To support bandwidth adaptation with these prioritized layers and SGs, the proposed method uses two-part procedures. First, the client predicts its future bandwidth (BW) with the channel prediction model using measured received signal strength indicator (RSSI) and current BW. Second, the server selects appropriate SGs among SVC bitstreams according to the received BW. Fig. 1 depicts the conceptual procedures.

Part 1: Since H.264 SVC consists of the video coding layer (VCL) and network abstraction layer (NAL), the server calculates the required bitrate for each SG by analyzing the NAL header information and compiles the results as a list. Given the *n* enhancement layers and *m* slice groups, the proposed method sets their priorities as shown in Fig. 2. When the server receives ABW from the client, it estimates total bitrate (ETB: the required bitrate for whole SGs). Once the ETB is lower than available bandwidth (ABW), it calculates ETB again after eliminating the lowest prioritized SG from the SG table. Thus, the server selects the SGs to be sent to the client by above recursive calculation.

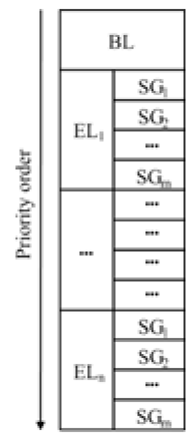


Figure 2: The prioritized SVC layers and slice groups.

Here, the unequal error protection (UEP) and FEC redundancy controlling method can be applied to the streaming policy using SG's priority. Since the required BW of each SG has a relatively large number, the FEC redundancy fills the remaining BW to the estimated ABW after SG selection to achieve fine granulate rate adaptation. This method is described in [4].

The flow chart in Fig. 3 describes the working processes of the server and the client.

Part 2: The client receives SG packets over a UDP channel from the server and then it measures the RSSI of a WIFI network. To maximize the performance of transmission, the client estimates the bandwidth not for time *t*, but for time *t* + α because the feedback has a delay α to be transmitted to the server. To get the estimated bandwidth in a WIFI network, this paper used the channel prediction model described in section 2 (it uses *t* + 1 instead of *t* + α because it uses α as a time unit).

2. Channel Prediction Model

This model explains how to predict the future RSSI value, say at time *t* + 1. Since the changes have a direction (drift: increment or decrement) when the client moves, the model predicts the future drift of the RSSI first.

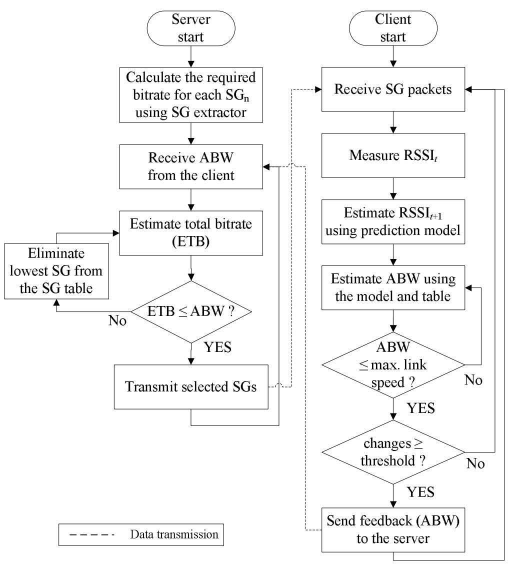


Figure 3: Channel prediction-based rate adaptation.

Prediction of the future drift of the RSSI: The realization of the drift can be represented by the first order difference

$$\Delta X_t = X_t - X_{t-1} \quad (1)$$

In order to estimate the future drift, we use the exponentially weighted moving average (EWMA) method, which was first proposed by Roberts [5], and widely used in *statistical process control chart*. For the theoretical detail, we refer to [6]. The predicted drift at time *t* + 1 is

$$\Delta\tilde{X}_{t+1} = \sum_{i=0}^{\infty} \gamma_i \Delta X_{t-i}, \quad (2)$$

where $\gamma_i = \alpha(1-\alpha)^i$ with $0 < \alpha \leq 1$.

The approach is to use higher weights on recent values and lower weights on past values. The coefficient α is considered a forgetting factor. Equation (2) can be rewritten in a recursive form as

$$\Delta\tilde{X}_{t+1} = (1-\alpha)\Delta\tilde{X}_t + \alpha\Delta X_t, \quad (3)$$

where $\Delta\tilde{X}_2 = \Delta X_2 = X_2 - X_1$.

Estimation of the current RSSI value: In the next step, we need to estimate the current RSSI by EWMA.

$$\hat{X}_t = (1-\beta)\hat{X}_{t-1} + \beta X_t, \quad (4)$$

where $0 < \beta \leq 1$.

Prediction of the future RSSI value: Based on the estimated RSSI $\hat{X}_t$ and the predicted drift $\Delta\tilde{X}_{t+1}$, we can predict the future RSSI in the following manner.

$$\tilde{X}_{t+1} = \hat{X}_t + \Delta\tilde{X}_{t+1} \quad (5)$$

Prediction of bandwidth from a sender: Let $\tilde{Y}_{t+1}$ be the predicted bandwidth from a sender with the decision rule for changing the ABW. In a WIFI network, the maximum link-speed (BW) is controlled by a pre-defined RSSI-link speed table. Suppose that we have M levels of the bandwidth and when the RSSI goes below the given threshold, K_p, the medium variable $\tilde{Y}_{t+1}$ of the bandwidth is reduced by d_p, which is written as

$$\tilde{Y}_{t+1} = B - \sum_{p=1}^{M} d_p I(\tilde{X}_{t+1} \leq K_p).$$

$I(\cdot)$ is an indicator function such that $I(\cdot) = 1$ when the condition is true, otherwise 0.

4. Conclusions

This paper proposes the SG-based SVC rate adaptation using the channel prediction model. It has several merits. First, it is a very light-weight adaptation method because the adaptation process is conducted not in an encoding step but in a transmission step. Second, it supports a more precise adaptation by estimating the bandwidth for the transmission time using the channel prediction model. This method can be applied to the STB, which supports in-home wireless video streaming to multiple clients.

References

[1] E-S Ryu and N Jayant, "Architecture of a home gateway for three-screen TV," Consumer Electronics (ICCE), 2011 IEEE International Conference on, pp.225-226, Jan. 2011.

[2] H Chen, Z Han, R Hu, and R Ruan, "Adaptive FMO selection strategy for error resilient H.264 coding," International Conference on Audio, Language and Image Processing (ICALIP 2008), pp. 868 - 872, Jul 2008.

[3] P Lambert, P Debevere, J De Cock et al., "Real-time

error concealing bitstream adaptation methods for SVC in IPTV systems," Journal of Real- Time Image Processing, Vol. 4, No. 1, pp. 79-90 Mar, 2009.
[4] Three Screen Television: The Home Gateway, URL:http://www.ece.gatech.edu/research/labs/MMC/research.html#three_screen
[5] S. W. Roberts, "Control chart tests based on geometric moving averages," Technometrics, Vol. 42, No. 1, pp. 97-101, Feb. 2000.
[6] M Basseville, IV Nikiforov, "Detection of abrupt changes - theory and application," Prentice-Hall: Englewoood Cliffs, N.J., 1993.

Eun Seok Ryu is a postdoctoral research fellow at Georgia Centers for Advanced Telecommunications Technology (GCATT) in Georgia Institute of Technology. Prior to joining GT, he was a research professor at Research Institute for Information and Communication Technology in Korea University. His research interests are in the area of multimedia communications, including joint source channel coding technologies and wireless mobile systems. He received his B.S., M.S., and Ph.D. in computer science from Korea University in 1999, 2001, and 2008 respectively.

Sung Won Han is a researcher in the Department of Biostatistics and Epidemiology at the University of Pennsylvania. He received an MS in Operations Research in 2006, an MS in Statistics in 2007, an MS in Mathematics in 2010, and a PhD in Industrial and Systems Engineering in 2010 from the Georgia Institute of Technology. His research interest is the sequential change-point detection problem in health care application and the statistical estimation in wireless network.

MMTC NEWS

3DRPC IG Conference Technical Call

Title: Selected Topics on the Kinect Sensor: Body Tracking and Facial Expression Tracking

Speaker: Cha Zhang, Microsoft Research, USA

Date/Time: Wednesday, June 22, 2011 at 7:00AM Pacific Time

Abstract: The Kinect sensor has attracted a lot of research interest recently. In this talk, we briefly present two works related to the Kinect sensor: how does the body tracking module work, and how to track facial expressions based on the Kinect sensor. For the first topic, we explain the underlying mechanism of body tracking in Kinect, which has enabled many popular games to use human body as the controller. For the second topic, we show how to use the depth and color information captured by the Kinect sensor to track facial expressions for human-computer interaction.

Bio: Cha Zhang is a Researcher in the Communication and Collaboration Systems Group at Microsoft Research (Redmond, WA). He received the B.S. and M.S. degrees from Tsinghua University, Beijing, China in 1998 and 2000, respectively, both in Electronic Engineering, and the Ph.D. degree in Electrical and Computer Engineering from Carnegie Mellon University, in 2004. His current research focuses on applying various machine learning and computer graphics/computer vision techniques to multimedia applications, in particular, multimedia teleconferencing. Dr. Zhang has published more than 40 technical papers and holds 10+ U.S. patents. He won the best paper award at ICME 2007, the top 10% award at MMSP 2009, and the best student paper award at ICME 2010. He co-authored two books titled "Light Field Sampling" and "Boosting-Based Face Detection and Adaptation", published by Morgan and Claypool in 2006 and 2010, respectively. Dr. Zhang is a Senior Member of IEEE. He was the Publicity Chair for International Packet Video Workshop in 2002, the program Co-Chair for the first Immersive Telecommunication Conference (IMMERSCOM) in 2007, the Steering Committee Co-Chair and Publicity Chair for IMMERSCOM 2009, the Program Co-Chair for the ACM Workshop on Media Data Integration (in conjunction with ACM Multimedia 2009), and the Poster&Demo Chair for ICME 2011. He served as TPC members for many conferences including ACM Multimedia, CVPR, ICCV, ECCV, MMSP, ICME, ICPR, ICWL, etc. He currently serves as an Associate Editor for Journal of Distance Education Technologies, IPSJ Transactions on Computer Vision and Applications, and ICST Transactions on Immersive Telecommunications. He was a guest editor for Advances in Multimedia, Special Issue on Multimedia Immersive Technologies and Networking.

CALL FOR PAPERS

IEEE Workshop on Multimedia Quality of Experience

Date: December 5-7, 2011

Place: Laguna Cliffs Marriott Resort & Spa, Dana Point, California USA

Web Link: http://www.dialogic.com/ieee/ism-december-2011/index.HTM

Organizers:
Tao Liu, Dialogic Research, Inc., USA
Junyong You, Norwegian University of Science and Technology, Norway