Michael J. Maher (Ed.)

Advances in Computer Science – ASIAN 2004

Higher-Level Decision Making

LNCS 3321

9th Asian Computing Science Conference
Dedicated to Jean-Louis Lassez on the Occasion of His 5th Cycle Birthday
Chiang Mai, Thailand, December 2004, Proceedings

Springer

Michael J. Maher (Ed.)

Advances in Computer Science - ASIAN 2004

Higher-Level Decision Making

9th Asian Computing Science Conference
Dedicated to Jean-Louis Lassez
on the Occasion of His 5th Cycle Birthday
Chiang Mai, Thailand, December 8-10, 2004
Proceedings

Springer

Volume Editor

Michael J. Maher
University of New South Wales
National ICT Australia, Sydney Laboratory
Locked Bag 6016, Sydney NSW 1466, Australia
E-mail: michael.maher@nicta.com.au

Library of Congress Control Number: Applied for

CR Subject Classification (1998): F.1-2, H.4, F.3, I.2, D.3, C.2.4, E.3, I.4, G.2

ISSN 0302-9743
ISBN 3-540-24087-X Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11355922 06/3142 5 4 3 2 1 0

Lecture Notes in Computer Science 3321

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

Lecture Notes in Computer Science

For information about Vols. 1–3225

please contact your bookseller or Springer

Vol. 3337: J.M. Barreiro, F. Martin-Sanchez, V. Maojo, F. Sanz (Eds.), Biological and Medical Data Analysis. XI, 508 pages. 2004.

Vol. 3333: K. Aizawa, Y. Nakamura, S. Satoh (Eds.), Advances in Multimedia Information Processing - PCM 2004, Part III. XXXV, 785 pages. 2004.

Vol. 3332: K. Aizawa, Y. Nakamura, S. Satoh (Eds.), Advances in Multimedia Information Processing - PCM 2004, Part II. XXXVI, 1051 pages. 2004.

Vol. 3331: K. Aizawa, Y. Nakamura, S. Satoh (Eds.), Advances in Multimedia Information Processing - PCM 2004, Part I. XXXVI, 667 pages. 2004.

Vol. 3323: G. Antoniou, H. Boley (Eds.), Rules and Rule Markup Languages for the Semantic Web. X, 215 pages. 2004.

Vol. 3322: R. Klette, J. Žunić (Eds.), Combinatorial Image Analysis. XII, 760 pages. 2004.

Vol. 3321: M.J. Maher (Ed.), Advances in Computer Science - ASIAN 2004. XII, 510 pages. 2004.

Vol. 3316: N.R. Pal, N.K. Kasabov, R.K. Mudi, S. Pal, S.K. Parui (Eds.), Neural Information Processing. XXX, 1368 pages. 2004.

Vol. 3315: C. Lemaître, C.A. Reyes, J.A. González (Eds.), Advances in Artificial Intelligence – IBERAMIA 2004. XX, 987 pages. 2004. (Subseries LNAI).

Vol. 3312: A.J. Hu, A.K. Martin (Eds.), Formal Methods in Computer-Aided Design. XI, 445 pages. 2004.

Vol. 3311: V. Roca, F. Rousseau (Eds.), Interactive Multimedia and Next Generation Networks. XIII, 287 pages. 2004.

Vol. 3309: C.-H. Chi, K.-Y. Lam (Eds.), Content Computing. XII, 510 pages. 2004.

Vol. 3308: J. Davies, W. Schulte, M. Barnett (Eds.), Formal Methods and Software Engineering. XIII, 500 pages. 2004.

Vol. 3307: C. Bussler, S.-k. Hong, W. Jun, R. Kaschek, Kinshuk, S. Krishnaswamy, S.W. Loke, D. Oberle, D. Richards, A. Sharma, Y. Sure, B. Thalheim (Eds.), Web Information Systems – WISE 2004 Workshops. XV, 277 pages. 2004.

Vol. 3306: X. Zhou, S. Su, M.P. Papazoglou, M.E. Orlowska, K.G. Jeffery (Eds.), Web Information Systems – WISE 2004. XVII, 745 pages. 2004.

Vol. 3305: P.M.A. Sloot, B. Chopard, A.G. Hoekstra (Eds.), Cellular Automata. XV, 883 pages. 2004.

Vol. 3303: J.A. López, E. Benfenati, W. Dubitzky (Eds.), Knowledge Exploration in Life Science Informatics. X, 249 pages. 2004. (Subseries LNAI).

Vol. 3302: W.-N. Chin (Ed.), Programming Languages and Systems. XIII, 453 pages. 2004.

Vol. 3299: F. Wang (Ed.), Automated Technology for Verification and Analysis. XII, 506 pages. 2004.

Vol. 3298: S.A. McIlraith, D. Plexousakis, F. van Harmelen (Eds.), The Semantic Web – ISWC 2004. XXI, 841 pages. 2004.

Vol. 3295: P. Markopoulos, B. Eggen, E. Aarts, J.L. Crowley (Eds.), Ambient Intelligence. XIII, 388 pages. 2004.

Vol. 3294: C.N. Dean, R.T. Boute (Eds.), Teaching Formal Methods. X, 249 pages. 2004.

Vol. 3293: C.-H. Chi, M. van Steen, C. Wills (Eds.), Web Content Caching and Distribution. IX, 283 pages. 2004.

Vol. 3292: R. Meersman, Z. Tari, A. Corsaro (Eds.), On the Move to Meaningful Internet Systems 2004: OTM 2004 Workshops. XXIII, 885 pages. 2004.

Vol. 3291: R. Meersman, Z. Tari (Eds.), On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE, Part II. XXV, 824 pages. 2004.

Vol. 3290: R. Meersman, Z. Tari (Eds.), On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE, Part I. XXV, 823 pages. 2004.

Vol. 3289: S. Wang, K. Tanaka, S. Zhou, T.W. Ling, J. Guan, D. Yang, F. Grandi, E. Mangina, I.-Y. Song, H.C. Mayr (Eds.), Conceptual Modeling for Advanced Application Domains. XXII, 692 pages. 2004.

Vol. 3288: P. Atzeni, W. Chu, H. Lu, S. Zhou, T.W. Ling (Eds.), Conceptual Modeling – ER 2004. XXI, 869 pages. 2004.

Vol. 3287: A. Sanfeliu, J.F. Martínez Trinidad, J.A. Carrasco Ochoa (Eds.), Progress in Pattern Recognition, Image Analysis and Applications. XVII, 703 pages. 2004.

Vol. 3286: G. Karsai, E. Visser (Eds.), Generative Programming and Component Engineering. XIII, 491 pages. 2004.

Vol. 3285: S. Manandhar, J. Austin, U.B. Desai, Y. Oyanagi, A. Talukder (Eds.), Applied Computing. XII, 334 pages. 2004.

Vol. 3284: A. Karmouch, L. Korba, E.R.M. Madeira (Eds.), Mobility Aware Technologies and Applications. XII, 382 pages. 2004.

Vol. 3283: F.A. Aagesen, C. Anutariya, V. Wuwongse (Eds.), Intelligence in Communication Systems. XIII, 327 pages. 2004.

Vol. 3282: V. Guruswami, List Decoding of Error-Correcting Codes. XIX, 350 pages. 2004.

Vol. 3281: T. Dingsøyr (Ed.), Software Process Improvement. X, 207 pages. 2004.

Vol. 3280: C. Aykanat, T. Dayar, İ. Körpeoğlu (Eds.), Computer and Information Sciences - ISCIS 2004. XVIII, 1009 pages. 2004.

Vol. 3278: A. Sahai, F. Wu (Eds.), Utility Computing. XI, 272 pages. 2004.

Vol. 3274: R. Guerraoui (Ed.), Distributed Computing. XIII, 465 pages. 2004.

Vol. 3273: T. Baar, A. Strohmeier, A. Moreira, S.J. Mellor (Eds.), <<UML>> 2004 - The Unified Modelling Language. XIII, 454 pages. 2004.

Vol. 3271: J. Vicente, D. Hutchison (Eds.), Management of Multimedia Networks and Services. XIII, 335 pages. 2004.

Vol. 3270: M. Jeckle, R. Kowalczyk, P. Braun (Eds.), Grid Services Engineering and Management. X, 165 pages. 2004.

Vol. 3269: J. Lopez, S. Qing, E. Okamoto (Eds.), Information and Communications Security. XI, 564 pages. 2004.

Vol. 3268: W. Lindner, M. Mesiti, C. Türker, Y. Tzitzikas, A. Vakali (Eds.), Current Trends in Database Technology - EDBT 2004 Workshops. XVIII, 608 pages. 2004.

Vol. 3266: J. Solé-Pareta, M. Smirnov, P.V. Mieghem, J. Domingo-Pascual, E. Monteiro, P. Reichl, B. Stiller, R.J. Gibbens (Eds.), Quality of Service in the Emerging Networking Panorama. XVI, 390 pages. 2004.

Vol. 3265: R.E. Frederking, K.B. Taylor (Eds.), Machine Translation: From Real Users to Research. XI, 392 pages. 2004. (Subseries LNAI).

Vol. 3264: G. Paliouras, Y. Sakakibara (Eds.), Grammatical Inference: Algorithms and Applications. XI, 291 pages. 2004. (Subseries LNAI).

Vol. 3263: M. Weske, P. Liggesmeyer (Eds.), Object-Oriented and Internet-Based Technologies. XII, 239 pages. 2004.

Vol. 3262: M.M. Freire, P. Chemouil, P. Lorenz, A. Gravey (Eds.), Universal Multiservice Networks. XIII, 556 pages. 2004.

Vol. 3261: T. Yakhno (Ed.), Advances in Information Systems. XIV, 617 pages. 2004.

Vol. 3260: I.G.M.M. Niemegeers, S.H. de Groot (Eds.), Personal Wireless Communications. XIV, 478 pages. 2004.

Vol. 3258: M. Wallace (Ed.), Principles and Practice of Constraint Programming – CP 2004. XVII, 822 pages. 2004.

Vol. 3257: E. Motta, N.R. Shadbolt, A. Stutt, N. Gibbins (Eds.), Engineering Knowledge in the Age of the Semantic Web. XVII, 517 pages. 2004. (Subseries LNAI).

Vol. 3256: H. Ehrig, G. Engels, F. Parisi-Presicce, G. Rozenberg (Eds.), Graph Transformations. XII, 451 pages. 2004.

Vol. 3255: A. Benczúr, J. Demetrovics, G. Gottlob (Eds.), Advances in Databases and Information Systems. XI, 423 pages. 2004.

Vol. 3254: E. Macii, V. Paliouras, O. Koufopavlou (Eds.), Integrated Circuit and System Design. XVI, 910 pages. 2004.

Vol. 3253: Y. Lakhnech, S. Yovine (Eds.), Formal Techniques, Modelling and Analysis of Timed and Fault-Tolerant Systems. X, 397 pages. 2004.

Vol. 3252: H. Jin, Y. Pan, N. Xiao, J. Sun (Eds.), Grid and Cooperative Computing - GCC 2004 Workshops. XVIII, 785 pages. 2004.

Vol. 3251: H. Jin, Y. Pan, N. Xiao, J. Sun (Eds.), Grid and Cooperative Computing - GCC 2004. XXII, 1025 pages. 2004.

Vol. 3250: L.-J. (LJ) Zhang, M. Jeckle (Eds.), Web Services. X, 301 pages. 2004.

Vol. 3249: B. Buchberger, J.A. Campbell (Eds.), Artificial Intelligence and Symbolic Computation. X, 285 pages. 2004. (Subseries LNAI).

Vol. 3246: A. Apostolico, M. Melucci (Eds.), String Processing and Information Retrieval. XIV, 332 pages. 2004.

Vol. 3245: E. Suzuki, S. Arikawa (Eds.), Discovery Science. XIV, 430 pages. 2004. (Subseries LNAI).

Vol. 3244: S. Ben-David, J. Case, A. Maruoka (Eds.), Algorithmic Learning Theory. XIV, 505 pages. 2004. (Subseries LNAI).

Vol. 3243: S. Leonardi (Ed.), Algorithms and Models for the Web-Graph. VIII, 189 pages. 2004.

Vol. 3242: X. Yao, E. Burke, J.A. Lozano, J. Smith, J.J. Merelo-Guervós, J.A. Bullinaria, J. Rowe, P. Tiňo, A. Kabán, H.-P. Schwefel (Eds.), Parallel Problem Solving from Nature - PPSN VIII. XX, 1185 pages. 2004.

Vol. 3241: D. Kranzlmüller, P. Kacsuk, J.J. Dongarra (Eds.), Recent Advances in Parallel Virtual Machine and Message Passing Interface. XIII, 452 pages. 2004.

Vol. 3240: I. Jonassen, J. Kim (Eds.), Algorithms in Bioinformatics. IX, 476 pages. 2004. (Subseries LNBI).

Vol. 3239: G. Nicosia, V. Cutello, P.J. Bentley, J. Timmis (Eds.), Artificial Immune Systems. XII, 444 pages. 2004.

Vol. 3238: S. Biundo, T. Frühwirth, G. Palm (Eds.), KI 2004: Advances in Artificial Intelligence. XI, 467 pages. 2004. (Subseries LNAI).

Vol. 3237: C. Peters, J. Gonzalo, M. Braschler, M. Kluck (Eds.), Comparative Evaluation of Multilingual Information Access Systems. XIV, 702 pages. 2004.

Vol. 3236: M. Núñez, Z. Maamar, F.L. Pelayo, K. Pousttchi, F. Rubio (Eds.), Applying Formal Methods: Testing, Performance, and M/E-Commerce. XI, 381 pages. 2004.

Vol. 3235: D. de Frutos-Escrig, M. Nunez (Eds.), Formal Techniques for Networked and Distributed Systems – FORTE 2004. X, 377 pages. 2004.

Vol. 3234: M.J. Egenhofer, C. Freksa, H.J. Miller (Eds.), Geographic Information Science. VIII, 345 pages. 2004.

Vol. 3233: K. Futatsugi, F. Mizoguchi, N. Yonezaki (Eds.), Software Security - Theories and Systems. X, 345 pages. 2004.

Vol. 3232: R. Heery, L. Lyon (Eds.), Research and Advanced Technology for Digital Libraries. XV, 528 pages. 2004.

Vol. 3231: H.-A. Jacobsen (Ed.), Middleware 2004. XV, 514 pages. 2004.

Vol. 3230: J.L. Vicedo, P. Martínez-Barco, R. Muñoz, M. Saiz Noeda (Eds.), Advances in Natural Language Processing. XII, 488 pages. 2004. (Subseries LNAI).

Vol. 3229: J.J. Alferes, J. Leite (Eds.), Logics in Artificial Intelligence. XIV, 744 pages. 2004. (Subseries LNAI).

Vol. 3226: M. Bouzeghoub, C. Goble, V. Kashyap, S. Spaccapietra (Eds.), Semantics of a Networked World. XIII, 326 pages. 2004.

Preface

The 9th Asian Computing Science Conference was held in Chiang Mai in December 2004. This volume contains papers that were presented at that conference. The conference was dedicated to Jean-Louis Lassez, on the occasion of his fifth cycle (60th year) birthday. Its theme was "higher-level decision-making".

Philippe Flajolet was invited to give the opening keynote address, while Yuzuru Tanaka and Phillip Rogaway were also keynote speakers. In addition to the keynote speakers, distinguished colleagues of Jean-Louis Lassez were invited to present talks in his honour. Many of those talks are represented by papers in this volume, but I would like to thank all those speakers:

Jonathan Bernick
Alex Brodsky
Vijay Chandru
T.Y. Chen
Norman Foo
Martin Golumbic
Richard Helm
Claude Kirchner
Hélène Kirchner
Kim Marriott
Krishna Palem
Kotagiri Ramamohanarao
Vijay Saraswat
Masahiko Sato
R.K. Shyamasundar
Nicolas Spyratos
Andrew Sung
Pascal Van Hentenryck
Jean Vuillemin

Following a call for papers, 75 papers were submitted, comprising 47 from Asia, 13 from Europe, 8 from Australia, 5 from North America, and 2 from the Middle East. Submissions came from a total of 23 countries. Korea, China and Thailand were the most-represented countries. These papers underwent anonymous refereeing before the Program Committee selected 17 papers for presentation at the conference. I thank the Program Committee for doing an excellent job under severe time pressure.

I thank the General Chairs of the conference, Joxan Jaffar and Kanchana Kanchanasut, for their support. I thank Pensri Arunwatanamongkol of AIT for her outstanding work in support of the conference website throughout the reviewing process. I thank the Local Organization Committee and its chair, Sanpawat

Kantabutra, for their work on the ground in Chiang Mai. Finally, I thank the Steering Committee for inviting me to chair the Program Committee.

I also recognize the sponsoring institutions for this conference:

Asian Institute of Technology, Thailand
Chiang Mai University, Thailand
IBM Research, USA
INRIA, France
National Electronics and Computer Technology Center, Thailand
National ICT Australia
National University of Singapore
Tata Institute of Fundamental Research, India
Waseda University, Japan

Michael Maher
Program Chair
ASIAN 2004

Preface

The 9th Asian Computing Science Conference was held in Chiang Mai in December 2004. This volume contains papers that were presented at that conference. The conference was dedicated to Jean-Louis Lassez, on the occasion of his fifth cycle (60th year) birthday. Its theme was "higher-level decision-making".

Philippe Flajolet was invited to give the opening keynote address, while Yuzuru Tanaka and Phillip Rogaway were also keynote speakers. In addition to the keynote speakers, distinguished colleagues of Jean-Louis Lassez were invited to present talks in his honour. Many of those talks are represented by papers in this volume, but I would like to thank all those speakers:

Jonathan Bernick
Alex Brodsky
Vijay Chandru
T.Y. Chen
Norman Foo
Martin Golumbic
Richard Helm
Claude Kirchner
Hélène Kirchner
Kim Marriott
Krishna Palem
Kotagiri Ramamohanarao
Vijay Saraswat
Masahiko Sato
R.K. Shyamasundar
Nicolas Spyratos
Andrew Sung
Pascal Van Hentenryck
Jean Vuillemin

Following a call for papers, 75 papers were submitted, comprising 47 from Asia, 13 from Europe, 8 from Australia, 5 from North America, and 2 from the Middle East. Submissions came from a total of 23 countries. Korea, China and Thailand were the most-represented countries. These papers underwent anonymous refereeing before the Program Committee selected 17 papers for presentation at the conference. I thank the Program Committee for doing an excellent job under severe time pressure.

I thank the General Chairs of the conference, Joxan Jaffar and Kanchana Kanchanasut, for their support. I thank Pensri Arunwatanamongkol of AIT for her outstanding work in support of the conference website throughout the reviewing process. I thank the Local Organization Committee and its chair, Sanpawat

Organization

Steering Committee

Joxan Jaffar (National University of Singapore, Singapore)
Gilles Kahn (INRIA, France)
Kanchana Kanchansut (Asian Institute of Technology, Thailand)
R.K. Shyamasundar (Tata Institute of Fundamental Research, India)
Kazunori Ueda (Waseda University, Japan)

Program Committee

Vijay Chandru (Strand Genomics and Indian Institute of Science, India)
T.Y. Chen (Swinburne University, Australia)
Philippe Codognet (Embassy of France in Tokyo and University of Paris 6, France)
Susumu Hayashi (Kobe University, Japan)
Jieh Hsiang (NTU, Taiwan)
Vitit Kantabutra (Idaho State University, USA)
Hélène Kirchner (LORIA-CNRS and INRIA, France)
Jimmy Lee (CUHK, Hong Kong, China)
Tze Yun Leong (NUS, Singapore)
Michael Maher (Loyola University Chicago, USA and NICTA, Australia)
Kim Marriott (Monash University, Australia)
Krishna Palem (Georgia Institute of Technology, USA)
Raghu Ramakrishnan (University of Wisconsin, Madison, USA)
Taisuke Sato (Tokyo Institute of Technology, Japan)
V.S. Subrahmanian (University of Maryland, College Park, USA)

Local Organization

Sanpawat Kantabutra (Chiang Mai University, Thailand)

Referees

David Austin
Steven Bird
Colin Boyd
Gerd Brewka
Christophe Cerisara
Lai-Wan Chan
Jeff Choi
Peter Chubb
Leonid Churilov
Johanne Cohen
Véronique Cortier
Christophe Doche
Alan Dorin
Spencer Fung
Claude Godart
Michael Goldwasser
Guido Governatori
Philippe de Groote
James Harland
L.C.K. Hui
Ryutaro Ichise
Yan Jin
Norman Foo
Waleed Kadous
Yoshitaka Kameya
Yukiyoshi Kameyama
Vellaisamy Kunalmani
Shonali Krishnaswam
François Lamarche
Phu Le
Dong Hoon Lee
Ho-Fung Leung
Guoliang Li
Lin Li
Jimmy Liu
Jim Lipton
Eric Martin
Ludovic Mé
Bernd Meyer
Thomas Meyer
Dale Miller
George Mohay
Yi Mu
Lee Naish
Amedeo Napoli
Nicolas Navet
Barry O'Sullivan
Maurice Pagnucco
Kuldip Paliwal
Shawn Parr
Silvio Ranise
Jochen Renz
Seiichiro Sakurai
Abdul Sattar
Jun Shen
John Shepherd
Arcot Sowmya
Peter Stuckey
Changai Sun
Willy Susilo
Antony Tang
Peter Tischer
Takehiro Tokuda
Mark Wallace
Kin-Hong Wong
Mariko Yasugi
Akihiro Yamamoto
Haruo Yokota
Jane You
Janson Zhang

Table of Contents

Keynote Papers

Contributed Papers

Invited Papers

Counting by Coin Tossings

Philippe Flajolet

Algorithms Project, INRIA-Rocquencourt, 78153 Le Chesnay, France
Philippe.Flajolet@inria.fr

Abstract. This text is an informal review of several randomized algorithms that have appeared over the past two decades and have proved instrumental in extracting efficiently quantitative characteristics of very large data sets. The algorithms are by nature probabilistic and based on hashing. They exploit properties of simple discrete probabilistic models and their design is tightly coupled with their analysis, itself often founded on methods from analytic combinatorics. Singularly efficient solutions have been found that defy information theoretic lower bounds applicable to deterministic algorithms. Characteristics like the total number of elements, cardinality (the number of distinct elements), frequency moments, as well as unbiased samples can be gathered with little loss of information and only a small probability of failure. The algorithms are applicable to traffic monitoring in networks, to data base query optimization, and to some of the basic tasks of data mining. They apply to massive data streams and in many cases require strictly minimal auxiliary storage.

1 Approximate Counting

Assume a blind man (a computer?) wants to locate a single black sheep amongst a flock of $N-1$ white sheep. The man can only ask an assistant questions with a Yes/No answer. Like for instance: "*Tell me whether the black sheep is amongst sheep ranked between 37 and 53 from the left*". This is a variation on the theme of "Twenty Questions". Clearly, about $\log_2 N \equiv \lg N$ operations are both necessary and sufficient. (The proof relies on the fact that with ℓ bits of information, you can only distinguish between 2^ℓ possibilities, so that one must have $2^\ell \geq N$, hence $\ell \geq \lg N$.) This simple argument is of an information-theoretic nature. It implies the fact that one cannot keep a counter (the "black sheep") capable of recording counts between 1 and N with less than $\lg N$ bits.

Assume that we want to run a counter known *a priori* to be in the range $1..N$. Can one beat the information-theoretic lower bound? Yes and No! Not if we require an exact count as this would contradict the information theoretic argument. But, ... Say we relax the constraints and tolerate an uncertainty on the counts of at most 10% (say) in relative terms. The situation changes dramatically. We now just need to locate our count amongst the terms of a geometric scale, $1, A, A^2, A^3 \ldots$ (till N), where $A = 1.1$. The problem then becomes that of

M.J. Maher (Ed.): ASIAN 2004, LNCS 3321, pp. 1–12, 2004.

finding an interval in a collection of about $\log_A N$ intervals. Information theory then tells us that this cannot be effected in fewer than

$$\lg \log_A N \approx \lg \lg N + 2.86245$$

bits, but it also tells us that the "amount of information" contained in an approximate answer is of that very same order. For instance, it is conceivable that an algorithm could exist with which counts till $N = 2^{16} \equiv 65536$ can be maintained using $8 + 3 = 11$ bits instead of 16.

This is the situation which Robert Morris encountered at Bell Labs in 1977. He needed to maintain the logs of a very large number of events in small registers, since the space available at the time was too small to allow exact counts to be kept. A gain by a factor of 2 in memory against a degradation in accuracy by some 30% was perfectly acceptable. How to proceed? Information theory provides a theoretical possibility, not a solution.

Morris's solution [23], known as the APPROXIMATE COUNTING Algorithm, goes at follows. In its crudest (binary) version, you maintain a counter K that initially receives the value $K := 0$. Counter K is meant to be a logarithmic counter, in the sense that when the exact count is n, the value K of the counter at that time should be close to $\lg n$. Note that the single value that gets stored is K, which itself only requires $\lg K \approx \lg \lg n$ bits. Morris' ingenious idea consists in updating the counter K to its new value $K^\star$ according to the following procedure:

$$K^\star = K + 1 \quad \text{with probability } 2^{-K}; \qquad K^\star = K \quad \text{with probability } 1 - 2^{-K}.$$

As time goes, the counter increases, but at a smaller and smaller pace. Implementation is especially easy given a (pseudo)random number generator of sorts [21].

The notable fact here is the appeal to a *probabilistic* idea in order to increment the counter. A plausible argument for the fact that K at time n should be close to $\lg n$ is the fact that it takes 1 impulse for the counter to go from value 0 to value 1, then on average 2 more impulses to from 1 to 2, then on average 4 more impulses from 2 to 3, and so on. In other words, a value $K = \kappa$ should be reached after about

$$1 + 2 + 4 + \cdots + 2^{\kappa - 1} = 2^\kappa - 1$$

steps. Thus, provided this informal reasoning is correct, one should have the numeric and probabilistic approximation $2^\kappa \approx n$. Then, the algorithm should return at each instant $n^\circ = 2^K$ as an estimate of the current value of n.

We have just exposed the binary version of the algorithm, which can at best provide an estimate within a factor of 2 since the values it returns are by design restricted to powers of 2. However, it is easy to change the *base* of the counter: it suffices to replace 2 by a smaller number q typically chosen of the form $q = 2^{1/r}$. Then, the new counter gets updated at basically r times the rate of the binary counter. Its granularity is improved, as is, we may hope, the accuracy of the result it provides.