Report from Dagstuhl Seminar 12481

Quantitative Security Analysis

Edited by

Boris Köpf[1], Pasquale Malacaria[2], and Catuscia Palamidessi[3]

1 **IMDEA Software Institute, ES**
2 **Queen Mary University of London, GB,** pm@dcs.qmw.ac.uk
3 **Ecole Polytechnique – Palaiseau, FR,** catuscia@lix.polytechnique.fr

Abstract

The high amount of trust put into today's software systems calls for a rigorous analysis of their security. Unfortunately, security is often in conflict with requirements on the functionality or the performance of a system, making perfect security an impossible or overly expensive goal. Under such constraints, the relevant question is not whether a system is secure, but rather how much security it provides. Quantitative notions of security can express degrees of protection and thus enable reasoning about the trade-off between security and conflicting requirements. Corresponding quantitative security analyses bear the potential of becoming an important tool for the rigorous development of practical systems, and a formal foundation for the management of security risks.

Seminar 20.–25. November, 2012 – www.dagstuhl.de/12481
1998 ACM Subject Classification C.2.0 General: Security and protection, E.4 Coding and Information Theory, D.3 Programming Languages, D.2.4 Software/Program Verification: Formal methods
Keywords and phrases Security, Privacy, Information theory, Programming languages, Formal methods
Digital Object Identifier 10.4230/DagRep.2.11.135

1 Executive Summary

Boris Köpf
Pasquale Malacaria
Catuscia Palamidessi

The high amount of trust put into today's software systems calls for a rigorous analysis of their security. Unfortunately, security is often in conflict with requirements on the functionality or the performance of a system, making perfect security an impossible or overly expensive goal. Under such constraints, the relevant question is not whether a system is secure, but rather how much security it provides. Quantitative notions of security can express degrees of protection and thus enable reasoning about the trade-off between security and conflicting requirements. Corresponding quantitative security analyses bear the potential of becoming an important tool for the rigorous development of practical systems, and a formal foundation for the management of security risks.

While there has been significant progress in research on quantitative notions of security and tools for their analysis and enforcement, existing solutions are still partial. The focus of the seminar is to discuss the following key issues.

Quantitative Security Analysis, *Dagstuhl Reports*, Vol. 2, Issue 11, pp. 135–154
Editors: Boris Köpf, Pasquale Malacaria, and Catuscia Palamidessi
DAGSTUHL REPORTS Dagstuhl Reports
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Quantitative Notions of Security: A single qualitative security property may give rise to a spectrum quantitative generalizations, each with different characteristics and application domains. For quantitative confidentiality, current research focuses on differential privacy and measures based on information-theoretic entropy. For other security properties such as integrity, availability, incoercibility, vote verifiability, etc., quantitative generalizations are only now emerging or have not even been proposed. One goal of this seminar is to advance the understanding of the relationship between existing quantitative security properties, and to join forces in the development of new ones.

Tools for Quantitative Security Analysis: Performing a quantitative security analysis of a realistic system is a challenging problem due to the complexity of modern software. It is mandatory to provide developers with tool support for this task. One goal of this seminar is to advance the understanding of the fundamental reasoning principles for quantitative notions of security, their connection to programming languages and verification techniques, and the theoretical limits for automatically deriving quantitative security guarantees.

Novel Application Domains: Quantitative security analyses have been successfully applied, e.g., for quantifying the side-channel leakage in cryptographic algorithms, for capturing the loss of privacy in statistical data analysis, and for quantifying security in anonymity networks. In emerging application domains such as electronic voting or distributed usage control, the need for quantitative analyses has been recognized. It is a goal of this seminar to foster the collaboration between experts in emerging application domains and those in quantitative security analysis.

2 Table of Contents

Executive Summary

Overview of Talks

3 Overview of Talks

3.1 Not all bits are created equal: incorporating the meaning and value of secret bits into measures of information flow

Mario Alvim (University of Pennsylvania, US)

Joint work of Alvim, Mário S.; Scedrov, Andre; Schneider, Fred B.

Most established information-theoretic approaches to quantitative information flow (QIF) define information in terms of Shannon entropy, min-entropy, or guessing entropy, which are a measure of, respectively, how much information flows, how likely it is that the secret be guessed in one try, and how long it takes to the secret to be guessed. These measures implicitly assume that every bit of the secret has the same "value", and therefore that every leaked bit represents the same 'threat". In many practical scenarios, however, some bits represent more important information than others (e.g. in a bank system, the bits representing the client's account number and pin code are more sensitive - and valuable – than the bits representing the client's street address). In this talk we discuss ongoing work on how to incorporate the "value" (i.e. meaning) of bits into measures of information and leakage. We consider deterministic systems, and use the Lattice of Information as an underlying algebraic structure for the set of all possible attacks an adaptive adversary can perform. We propose several measures for the information carried by the elements in the lattice. In particular, we are able to show that the measures for QIF based on Shannon entropy, min-entropy and guessing entropy are a special case of our approach, where every field of the secret is considered to be equally valuable.

3.2 Multi-run security

Arnar Birgisson (Chalmers UT – Göteborg, SE)

Joint work of Birgisson, Arnar; Sabelfeld, Andrei
Main reference A. Birgisson, A. Sabelfeld, "Multi-run security," in Proc. of the European Symp. on Research in Computer Security (ESORICS'11), LNCS, Vol. 6879, pp. 372–391, Springer, 2011.
URL http://dx.doi.org/10.1007/978-3-642-23822-2_21
URL http://www.hvergi.net/arnar/publications/pdf/esorics11.pdf

This paper explores information-flow control for batch-job programs that are allowed to be re-run with new input provided by the attacker. We argue that directly adapting two major security definitions for batch-job programs, termination-sensitive and termination-insensitive noninterference, to multi-run execution would result in extremes. While the former readily scales up to multiple runs, its enforcement is typically over-restrictive. The latter suffers from insecurity: secrets can be leaked in their entirety by multiple runs of programs that are secure according to batch-job termination-insensitive noninterference.

Seeking to avoid the extremes, we present a framework for specifying and enforcing multi-run security in an imperative language. The policy framework is based on tracking the attacker's knowledge about secrets obtained by multiple program runs. Inspired by previous work on robustness, the key ingredient of our type-based enforcement for multi-run security is preventing the dangerous combination of attacker-controlled data and secret data from affecting program termination.

3.3 Worst- and average-case privacy breaches in randomization mechanisms

Michele Boreale (Università di Firenze, IT)

Joint work of Boreale, Michele; Paolini, Michela
Main reference M. Boreale, M. Paolini, "Worst- and average-case privacy breaches in randomization mechanisms," in Proc. of IFIP-TCS 2012, LNCS, Vol. 7604, pp. 72–86, Springer, 2012.
URL http://dx.doi.org/10.1007/978-3-642-33475-7_6

In a variety of contexts, randomization is regarded as an effective technique to conceal sensitive information. We model randomization mechanisms as information-theoretic channels. Our starting point is a semantic notion of security that expresses absence of any privacy breach above a given level of seriousness ϵ, irrespective of any background information, represented as a prior probability on the secret inputs. We first examine this notion according to two dimensions: worst vs. average case, single vs. repeated observations. In each case, we characterize the security level achievable by a mechanism in a simple fashion that only depends on the channel matrix, and specifically on certain measures of "distance" between its rows, like norm-1 distance and Chernoff Information. We next clarify the relation between our worst-case security notion and differential privacy (dp): we show that, while the former is in general stronger, the two coincide if one confines to background information that can be factorised into the product of independent priors over individuals. We finally turn our attention to expected utility, in the sense of Ghosh et al., in the case of repeated independent observations. We characterize the exponential growth rate of any reasonable utility function. In the particular case the mechanism provides ϵ-dp, we study the relation of the utility rate with ϵ: we offer either exact expressions or upper-bounds for utility rate that apply to practically interesting cases, such as the (truncated) geometric mechanism.

3.4 Measuring Information Leakage using Generalized Gain Functions

Kostas Chatzikokolakis (Ecole Polytechnique – Palaiseau, FR)

Joint work of Alvim, Mario; Chatzikokolakis, Kostas; Palamidessi, Catuscia; Smith, Geoffrey
Main reference M.S. Alvim, K. Chatzikokolakis, C. Palamidessi, G. Smith, "Measuring Information Leakage using Generalized Gain Functions," in Proc. of the 25th IEEE Computer Security Foundations Symposium (CSF'12), pp. 265–279, IEEE CS, 2012.
URL http://dx.doi.org/10.1109/CSF.2012.26
URL http://www.lix.polytechnique.fr/~kostas/papers/csf12_gleakage.pdf

This talk introduces g-leakage, a rich generalization of the min-entropy model of quantitative information flow. In g-leakage, the benefit that an adversary derives from a certain guess about a secret is specified using a gain function g. Gain functions allow a wide variety of operational scenarios to be modeled, including those where the adversary benefits from guessing a value close to the secret, guessing a part of the secret, guessing a property of the secret, or guessing the secret within some number of tries. I will discuss important properties of g-leakage, including bounds between min-capacity, g-capacity, and Shannon capacity. Moreover I will discuss a connection between a strong leakage ordering on two channels, C_1 and C_2, and the possibility of factoring C_1 into C_2C_3, for some C_3. Based on this connection, I will propose a generalization of the Lattice of Information from deterministic to probabilistic channels.

3.5 Estimating Information Leakage from Trial Runs and Whole Java Programs.

Tom Chothia (University of Birmingham, GB)

In this talk I will outline some results that make it possible to estimate measures of information leakage based on mutual information and min-entropy from trial runs of a system alone. We propose statistical estimation as a method of applying more theoretical work on quantitative security directly to implemented systems, and we will demonstrate this by measuring the information leak in MIX nodes and from encrypted Tor traffic. We will then present a model of leakage for complete, probabilistic, non-terminating programs and show how we can use this to estimate the information leakage from large Java Programs.

3.6 Approximation and Relative Entropy

Alessandra Di Pierro (Università degli Studi di Verona, IT)

Joint work of Di Pierro, Alessandra; Hankin, Chris; Wiklicky, Herbert

Program analysis produces approximated results due to the abstraction on the state space which is required to construct 'simplified' (computable) semantics. In the case of probabilistic abstraction, i.e. when the abstract domain is a probability space, this approximation can be seen as the 'inefficiency' of mistakenly assuming that the behaviour of a source program P is a distribution x when the true distribution is y. In terms of information theory this is represented by the notion of relative entropy, aka Kullback-Leibler divergence [5]. Based on this intuition, we re-visit the notion of Approximate Confinement introduced in [4, 3, 2]. This notion formalises probabilistic non-interference in terms of process indistinguishability according to some abstract semantics (I/O observables, bisimulation etc.) and allows for the leakage of a certain amount epsilon of information. Such a quantity corresponds to a measure of the approximation introduced by the abstract semantics (probabilistic observables) and can thus be interpreted as a measure of the KL-divergence of the system. The statistical interpretation [1] of epsilon as an estimate of the number of tests needed to differentiate two executions of a program on sensitive data (i.e. how hard an attacker has to work in order to breach security) is also in accordance with the hypothesis testing formulation of relative entropy.

References

1 J. Shao, Mathematical Statistics, Springer Texts in Statistics, Springer Verlag, New York – Berlin – Heidelberg, 1999.

2 A. Di Pierro, C. Hankin, H. Wiklicky, Measuring the confinement of probabilistic systems, Theoretical Computer Science 340 (1) (June 2005) 3–56

3 A. Di Pierro, C. Hankin, H. Wiklicky, Approximate Non-Interference, Journal of Computer Security 12 (1) (2004) 37–81.

4 A. Di Pierro, C. Hankin, H. Wiklicky, Approximate non-interference, in: Proceedings of CSFW'02, IEEE, Cape Breton, Canada, 2002, pp. 3–17.

5 S. Kullback, R. A. Leibler, On Information and Sufficiency, Ann. Math. Statist. 22 (1) (1951) 79-86.

3.7 A differentially private mechanism of optimal utility for a region of priors

Ehab ElSalamouny (Ecole Polytechnique – Palaiseau, FR)

Joint work of ElSalamouny, Ehab; Chatzikokolakis, Konstantinos; Palamidessi, Catuscia
Main reference E. ElSalamouny, K. Chatzikokolakis, C. Palamidessi, "A differentially private mechanism of optimal utility for a region of priors," in Proc. of the 2nd Int'l Conf. on Principles of Security and Trust (POST'13), LNCS, Vol. 7796, pp. 41–62, Springer, 2013.
URL http://dx.doi.org/10.1007/978-3-642-36830-1_3
URL http://hal.inria.fr/docs/00/77/20/99/PDF/main.pdf

The notion of differential privacy has emerged in the area of statistical databases as a measure of protection of the participants sensitive information, which can be compromised by selected queries. In this talk I consider mechanisms which satisfy differential privacy by perturbing query outputs, and therefore reduce their utility. Since for any non-counting query there is no such a mechanism that is optimal for every prior (side knowledge), I highlight for an arbitrary query and a privacy parameter a special region of priors for which an optimal mechanism may exist. For each prior in this region, I show upper bounds for utility as well as for min-mutual information between the real query results and the noisy outputs reported to the user. Then, I describe a special mechanism, called the "tight- constraints mechanism", and discuss the conditions for its existence. This mechanism has the property of reaching the bounds for all the priors of the region, and thus it is optimal on the whole region. Finally I show that the same analysis implies tight upper-bounds for the min-entropy leakage about the database through any differentially private mechanism.

3.8 Quantifying Leakage in the Presence of Unreliable Source of Information

Sardaouna Hamadou (Ecole Polytechnique – Palaiseau, FR)

The frequent inaccurate, misleading or outdated material about people and businesses on social networks, online forums, blogs and other forms of online communication and information sharing raises important reputation and privacy issues.

In this talk, we will address these issues by providing a formal framework generalizing current methods of quantifying information flow. We will refine these models by integrating the notion of belief. The idea is that the threat should be relative to the possible initial belief, that is a (potentially inaccurate) information, that an attacker may have about the confidential information. More precisely, we will consider the case where the adversary is combining information from an external and potentially unreliable source and the observables of a program/protocol in order to increase her chance of breaking the privacy. In such context, concepts from Belief Theory have proved quite useful as it has higher ability to combine information from (partially or totally) disagreeing sources.

3.9 Attack Time Analysis

Holger Hermanns (Universität des Saarlandes, DE)

Joint work of Hermanns, Holger; Pulungan, Reza; Stoelinga, Marielle; Arnold, Florian

Security attacks are a threat to an increasing number of systems on which our society depends. Coined by Bruce Schneier, attack trees are a convenient graphical formalism to structure the understanding of potential security attacks and to quantify security risks.

An important concern in quantitative risk analysis is the quality of the data: how realistic are the probabilities attached to basic attack steps? In this work, we show how fitting techniques for phase type distributions can help in a time dependent risk analysis. The approach we propose combines this information with the attack tree, and turns it into a Markov chain. Compositional compression techniques are used to keep the size of this model manageable. The quantitative evaluation of this model is delegated to a stochastic model checker.

We apply this approach to a genuine attack tree example. This example reveals obvious further operators that seem natural to be included in an attack tree formalism.

3.10 Dynamic enforcement of knowledge-based security policies using probabilistic abstract interpretation

Michael Hicks (University of Maryland – College Park, US)

Joint work of Hicks, Michael; Mardziel, Piotr; Magill, Stephen; Srivatsa, Mudhakar
Main reference P. Mardziel, S. Magill, M. Hicks, M. Srivatsa, "Dynamic Enforcement of Knowledge-based Security Policies using Probabilistic Abstract Interpretation," (extended version of CSF'11 paper), October, 2012.
URL http://www.cs.umd.edu/~mwh/papers/mardziel12belieflong.html

This work explores the idea of knowledge-based security policies, which are used to decide whether to answer queries over secret data based on an estimation of the querier's (possibly increased) knowledge given the results. Limiting knowledge is the goal of existing information release policies that employ mechanisms such as noising, anonymization, and redaction. Knowledge-based policies are more general: they increase flexibility by not fixing the means to restrict information flow.

We enforce a knowledge-based policy by explicitly tracking a model of a querier's belief about secret data, represented as a probability distribution, and denying any query that could increase knowledge above a given threshold. We implement query analysis and belief tracking via abstract interpretation, which allows us to trade off precision and performance through the use of abstraction, while maintaining soundness. We have developed an approach to augment standard abstract domains to include probabilities, and thus define distributions. We focus on developing probabilistic polyhedra in particular, to support numeric programs. While probabilistic abstract interpretation has been considered before, our domain is the first whose design supports sound conditioning, which is required to ensure that estimates of a querier's knowledge are accurate.

Experiments with our implementation show that several useful queries can be handled efficiently, particularly compared to exact (i.e., sound) inference involving sampling. We also show that, for our benchmarks, restricting constraints to octagons or intervals, rather than full polyhedra, can dramatically improve performance while incurring little to no loss in precision.

Finally, I will sketch a generalization of our ideas to reasoning about information release in secure multiparty computations.

3.11 A Framework for Extracting Semantic Guarantees from Privacy Definitions

Daniel Kifer (Penn State University – University Park, US)

Joint work of Lin, Bing-Rong; Kifer, Daniel
Main reference B.-R. Lin, D. Kifer, "A Framework for Extracting Semantic Guarantees from Privacy," arXiv:1208.5443v1 [cs.DB].
URL http://arxiv.org/abs/1208.5443

The goal of statistical privacy is to choose an algorithm whose input is a sensitive data set and whose output contains useful and nonsensitive statistical information.

Privacy definitions specify the algorithms that can be used. However there is a long history of finding new weaknesses in existing privacy definitions. Thus there is a need for tools for analyzing privacy definitions. One common approach is to invent an attack and see if certain pieces of information can be inferred. This is hit-or-miss: if an attack does not work then no conclusions can be drawn, and it will not identify cases where the privacy definition protects unnecessary pieces of information.

In this talk we present a new framework for analyzing privacy definitions and deriving their semantic guarantees.

3.12 How We Should Measure the Utility of Sanitizing Mechanisms, How We Should Process Sanitized Data, and Why.

Daniel Kifer (Penn State University – University Park, US)

Joint work of Lin, Bing-Rong; Kifer, Daniel

Utility measures are used in statistical privacy to identify algorithms that produce the most useful output subject to privacy constraints. In this talk we analyze utility measures that are commonly used in the literature and show that they reward algorithms for throwing away information.

To fix this problem, we consider 3 axioms that utility measures should satisfy. Surprisingly, these axioms imply that utility should be measured using Bayesian Decision Theory even though there is nothing Bayesian about those axioms (in fact, they are quite different from the usual derivations of Bayesian decision theory).

These results imply that choosing sanitizing algorithms to maximize utility is the same as choosing sanitizing algorithms whose outputs are best analyzed using decision theoretical tools (as opposed to other statistical procedures). We conduct an experimental evaluation on the privacy-preserving sorted histogram problem and empirically show that the decision-theoretic tools consistently produce more accurate estimators than previous approaches.

3.13 Automatic Quantification of Cache Side Channels

Boris Köpf (IMDEA Software Institute, ES)

Joint work of Doychev, Goran; Feld, Dominik; Köpf, Boris; Mauborgne, Laurent; Ochoa, Martin; Reineke, Jan
Main reference B. Köpf, L. Mauborgne, M. Ochoa, "Automatic Quantification of Cache Side Channels," in Proc. of the 24th Int'l Conf. on Computer-Aided Verification (CAV'12), LNCS, Vol. 7358, pp. 564–580, Springer, 2012.
URL http://dx.doi.org/10.1007/978-3-642-31424-7_40

This talk presents work on a novel approach for automatically deriving upper bounds on the amount of information about the input that an adversary can extract from a program by observing the CPU's cache behavior.

Technically, our approach builds on the observation that (an upper bound on) the number of possible side-channel observations corresponds to (an upper bound on) the number of leaked bits. Such upper bounds can be obtained by computing super-sets of the set of possible observations by abstract interpretation, and by determining sizes [3]. We apply this idea to the problem of quantifying cache side-channels by providing abstract domains that keep track of the side-channel observations different kinds of cache adversaries can make, together with counting procedures for the number of corresponding concretisations.

The first part of this talk presents an initial case study, in which we combine existing tools for static cache analysis [2, 1] with a novel counting procedure for cache states. We use this combination for deriving bounds on the leakage of executables of standard AES implementations, demonstrating that automatically deriving security guarantees against cache attacks is indeed feasible. However, the obtained bounds hold only for a certain class of adversaries (namely: access-based), and their derivation requires code instrumentation.

The second part of this talk presents ongoing work on a dedicated tool for the automatic quantification of cache side-channels. The tool is based on an abstract interpretation engine for x86 binaries and can be easily extended by abstract domains for different kinds of cache observations. We provide a novel set of such abstract domains that cover all kinds of adversary models that are typically considered in the literature, namely: access-based, trace-based, and time-based. The talk concludes with experimental results, including the first security proof of the preloading countermeasure, based on an actual x86 executable of AES.

Talk is based on joint work with Goran Doychev, Dominik Feld, Laurent Mauborgne, Martin Ochoa, and Jan Reineke.

References

1 AbsInt aiT Worst-Case Execution Time Analyzers. http://www.absint.com/a3/

2 C. Ferdinand, F. Martin, R. Wilhelm, and M. Alt. Cache behavior prediction by abstract interpretation. *Science of Computer Programming*, 35(2):163 – 189, 1999.

3 B. Köpf and A. Rybalchenko. Approximation and Randomization for Quantitative Information-Flow Analysis. In *CSF*, pages 3–14. IEEE, 2010.

3.14 Algebraic Foundations for Quantitative Information Flow

Pasquale Malacaria (Queen Mary University of London, GB)

Several mathematical ideas have been investigated for Quantitative Information Flow. Information theory, probability, guessability are the main ideas in most proposals. They aim to quantify *how much information* is leaked, *how likely is to guess* the secret and *how long does it take* to guess the secret respectively. In this work we investigate the relationship between these ideas in the context of the quantitative analysis of deterministic systems. We propose the Lattice of Information as a valuable foundation for these approaches; not only it provides an elegant algebraic framework for the ideas, but also to investigate their relationship. In particular we will use this lattice to prove some results establishing order relation correspondences between the different quantitative approaches. The implications of these results w.r.t. recent work in the community is also investigated.

While this work concentrates on the foundational importance of the Lattice of Information its practical relevance has been recently proven, notably with the quantitative analysis of Linux kernel vulnerabilities. Overall we believe these works set the case for establishing the Lattice of Information as one of the main reference structure for Quantitative Information Flow.

The talk is based on the forthcoming paper [1].

References

1 P. Malacaria. Algebraic Foundations for Quantitative Information Flow. Mathematical Structures in Computer Science (To appear)

3.15 Denotational models for non-interference, probability and nondeterminism

C. Carroll Morgan (UNSW – Sydney, AU)

Joint work of McIver, Annabelle K.; Meinicke, Larissa A.; Morgan, C. Carroll
Main reference A. McIver, L. Meinicke, C. Morgan, “Compositional closure for Bayes Risk in probabilistic noninterference,” in Proc. of the 37th Int’l Colloquium on Automata, Languages and Programming – Part II (ICALP’10), LNCS, Vol. 6199, pp. 223–235, Springer, 2010.
URL http://dx.doi.org/10.1007/978-3-642-14162-1_19

Combining the three features of the title is a notoriously hard problem that has attracted much interesting work. Actually the fourth feature is "denotational", as producing a denotational semantics (rather than only a mathematical model) introduces novel and challenging further constriants.

Markov Processes (for probability without nondeterminism or hiding), Markov Decision Processes (add nondeterminism) and Partially Observable MDP's (add hiding) are mathematical models. Adding the further constraints of full abstraction and compositionality, encouraged by a computer-science perspective, suggest quotients on these models that induce interesting semantic domains built from monads and that include novel "security refinement" partial orders.

I will describe the above motivations, briefly, and then in more detail describe key features of the models we have constructed with all those constraints in mind. The target for the talk will be to explain the steps we are taking with our most recent work, and to elicit suggestions from the group about possible ways forward.

As part of our contribution we proved the outstanding "Coriaceous Conjecture", an open problem due to Alvim, Chatzikokolakis, Palamidessi and Smith (also participants at the seminar). See the Seminar-Wide materials section.

The work is joint, with Annabelle McIver (Macquarie University) and Larissa Meinicke (University of Queensland).

References

1 McIver, AK, Meinicke, LA, Morgan CC. *Compositional closure for Bayes Risk in probabilistic noninterference.* Proc. ICALP 2010. http://dx.doi.org/10.1007/978-3-642-14162-1_19

2 Morgan, CC. *Compositional noninterference from first principles.* Formal Aspects of Computing 24(1):1-24. Springer (2010) http://dx.doi.org/10.1007/s00165-010-0167-y

3 McIver, AK, Meinicke, LA, Morgan CC. *A Kantorovich-Monadic Powerdomain for Information Hiding, with Probability and Nondeterminism.* Proc. LiCS 2012. IEEE (2012). http://dx.doi.org/10.1109/LICS.2012.56

3.16 Probabilistic model checking and PRISM

Gethin Norman (University of Glasgow, GB)

Main reference M. Kwiatkowska, G. Norman, D. Parker, "PRISM 4.0: Verification of Probabilistic Real-time Systems," in Proc. of 23rd Int'l Conf. on Computer Aided Verification (CAV'11), LNCs, Vol. 6806, pp. 585–591, Springer, 2011.
URL http://dx.doi.org/10.1007/978-3-642-22110-1_47

Probabilistic model checking has established itself as a valuable technique for the formal modelling and analysis of systems that exhibit stochastic behaviour. It has been applied to of a wide range of systems, from communication and security protocols to biological signalling pathways. This talk gives an overview of my research in this area, reviewing the different the models, properties and application domains that have been investigated. In addition, I will outline some techniques being developed to allow for the efficient analysis of complex and infinite state systems, as well as some open problems and future challenges.

3.17 Indistinguishable regions in Geographic Privacy

Martin Ochoa (Siemens – München, DE)

Joint work of Cuellar, Jorge; Ochoa, Martin; Rios, Ruben
Main reference J. Cuellar, M. Ochoa, R. Rios, "Indistinguishable regions in Geographic Privacy," in Proc. of the 27th Annual ACM Symp. on Applied Computing (SAC'12), pp. 1463–1469, ACM, 2012.
URL http://dx.doi.org/10.1145/2245276.2232010

The ubiquity of positioning devices poses a natural security challenge: users want to take advantage of location-related services as well as social sharing of their position but at the same time have security concerns about how much information should be shared about their exact position. This talk discusses different location-privacy problems, their formalization and the novel notion of indistinguishability regions that allows one to proof that a given obfuscation function provides a good trade-off between location sharing and privacy.

3.18 Enhancing Differential Privacy: from Hamming to General Metrics

Catuscia Palamidessi (Ecole Polytechnique – Palaiseau, FR)

Differential Privacy is one of the most prominent frameworks used to deal with disclosure prevention in statistical databases. Differential privacy is a formal privacy guarantee that ensures that sensitive information relative to individuals cannot be easily inferred by disclosing answers to aggregate queries. If two databases are adjacent, i.e. differ only for an individual, then querying them should not allow to tell them apart by more than a certain factor. The transitive application of this property induces a bound also on the distinguishability of two generic databases, which is determined by their distance on the Hamming graph of the adjacency relation.

In this paper we lift the restriction relative to the Hamming graphs and we explore the implications of differential privacy when the indistinguishability requirement depends on an arbitrary notion of distance. We show that we can express, in this way, (protection against) kinds of privacy threats that cannot be naturally represented with the standard notion. We give an intuitive characterization of these threats in terms of Bayesian adversaries, which generalizes the characterization of (standard) differential privacy from the literature. Next, we revisit the well-known result on the non-existence of universally optimal mechanisms for any query other than counting queries. We show that in our setting, for certain kinds of distances, there are many more queries for which universally optimal mechanisms exist: Notably sum, average, and percentile queries. Finally, we show some applications in various domains: statistical databases where the units of protection are groups (rather than individuals), geolocation, and smart metering.

3.19 Towards an SMT-based approach for Quantitative Information Flow

Quoc Sang Phan (Queen Mary University of London, GB)

Joint work of Phan, Quoc Sang; Malacaria, Pasquale

Quantitative Information Flow (QIF) is a powerful approach to analyse leaks of confidential information in a software system. Here we present a novel method for automated QIF analysis. We cast the problem of QIF analysis into the problem of counting boolean abstractions of satisfiable instances of a hidden SMT formula. We present a DPLL($\mathcal{T}$)-like framework to build a solver for this problem. We then prove that the methodology of Symbolic Execution also fits our framework. Based on these ideas, we build two QIF analysis tools: the first one employs CBMC, a bounded model checker for ANSI C, and the second one is built on top of Symbolic Pathfinder, a Symbolic Execution tool for Java. For experiment, we quantify information leakage of programs in the Linux kernel, and analyse a tax program in Java taken from the European project HATS.

3.20 Quantitative Distributed Data Usage Control

Alexander Pretschner (TU München, DE)

Joint work of Pretschner, Alexander; Lovat, Enrico; Oudinet, Johan

Distributed data usage control is about what happens to data once it is given away ("delete after 30 days;" "notify me if data is forwarded;" "copy at most twice"). In the past, we have considered the problem in terms of policies, enforcement and guarantees from two perspectives: (a) In order to protect data, it is necessary to distinguish between content (a song by Elvis called "Love me Tender") and representations of that content (song.mp3; song.wav, etc.). This requires data flow-tracking concepts and capabilities in data usage control frameworks. (b) These representations exist at different layers of abstraction: a picture downloaded from the internet exists as pixmap (window manager), as element in the browser-created DOM tree (application), and as cache file (operating system). This requires the data flow tracking capabilities to transcend the single layers to which they are deployed. In distributed systems, it has turned out that another system can be seen as another set of abstraction layers, thus generalizing the basic model. Demo videos of this work available at http://www22.in.tum.de/forschung/distributed-usage-control/.

In this talk, we present recent work on extending our approach to not only protecting entire data items but possibly also fractions of data items. This allows us to specify and enforce policies such as "not more than 20% of the data may leave the system", evidently leading to interesting questions concerning the interpretation of "20%", and if the structure of data items cannot be exploited. We present a model, its instantiation to the operating system layer, and first experimental results.

3.21 Constraint Solving Based on Horn Clauses for Verifying Information Flow Properties

Andrey Rybalchenko (TU München, DE)

Joint work of Gleissenthall, Klaus; Köpf, Boris

We present a constraint generation and solving methods that can be used as building blocks for automatic verification of information flow properties of programs.

3.22 Quantifying Opacity

Mathieu Sassolas (Université Libre de Bruxelles, BE)

Joint work of Bérard, Béatrice; Mullins, John; Sassolas, Mathieu
Main reference B. Bérard, J. Mullins, M. Sassolas, "Quantifying Opacity," arXiv:1301.6799v2 [cs.CR]. To appear in Mathematical Structures in Computer Science, Cambridge University Press.
URL http://arxiv.org/abs/1301.6799

Opacity is a general language-theoretic scheme [1] which can be instanciated into several security properties of a system. Its parameters are a predicate, given as a subset of runs of the system, and an observation function, from the set of runs into a set of observables. The predicate describes secret information in the system and, in the possibilistic setting, it is opaque if its membership cannot be inferred from observation.

In this presentation, we propose several notions of quantitative opacity for probabilistic systems, where the predicate and the observation function are seen as random variables. The distribution of these variables is based on the distribution of the potentially infinite set of runs, and therefore is not given in an extensive form.

Our aim is to measure (i) the probability of opacity leakage relative to these random variables and (ii) the level of uncertainty about membership of the predicate inferred from observation. We show how these measures extend possibilistic opacity, we give algorithms to compute them for regular secrets and observations, and we apply these computations on the classical example of Crowds protocol.

As part of ongoing work, we also study approximate computation of these measures and the non-deterministic setting.

This talk is based on joint work with Béatrice Bérard and John Mullins [2].

References

1 Bryans, J.W., Koutny, M., Mazaré, L., Ryan, P.Y.A.: Opacity generalised to transition systems. Intl. Jour. of Information Security **7**(6) (2008) 421–435

2 Bérard, B., Mullins, J., Sassolas, M.: Quantifying opacity. Mathematical Structures in Computer Science (To appear)

3.23 Thermodynamic aspects of confidentiality: timing channels in Brownian computers

Fabrizio Smeraldi (Queen Mary University of London, GB)

Joint work of Malacaria, Pasquale; Smeraldi, Fabrizio

Timing channels are contingent both on the specific computation and on the architecture of the system - notably, the presence of a clock signal. Brownian computers are unclocked devices that operate through the combined action of thermal agitation and a weak driving potential (think computing with DNA). Under these conditions variations of the entropy of the computer due to, for instance, logically irreversible computations generally affect the dynamics of the system.

As a consequence, an entirely new category of timing channels emerges that allows discriminating between computations requiring the same number of steps on the basis of their different degree of irreversibility.

The talk is based on a forthcoming paper [1].

References

1 P. Malacaria, F. Smeraldi.: Thermodynamic Aspects of Confidentiality. Information and Computation (To appear).

3.24 Channels and Composition Refinement

Geoffrey Smith (Florida Int. Univ. – Miami, US)

Joint work of Espinoza, Barbara

Given channels C1 and C2 from a set X of secret inputs, it may be that C1 is equivalent to C2 followed by some post-processing; that is, C1 can be factored into the cascade of C2 and C3 for some channel C3. In this case we say that C1 is composition refined by C2. Composition refinement coincides with partition refinement in the Lattice of Information in the case when C1 and C2 are deterministic channels, but composition refinement is meaningful for probabilistic channels as well. In this talk, I discuss some current work on the mathematical structure of channels under the composition refinement relation, considering in particular the case when channels C1 and C2 are composition equivalent, meaning that each composition refines the other. I show that composition refinement is a partial order up to semantic equivalence, where channels are semantically equivalent if they are equal as maps from priors to hyper- distributions. I also mention some connections to Quantitative Information Flow.

The talk is based on joint work with Barbara Espinoza.

3.25 On Complexity of Verifying Quantitative Information Flow

Tachio Terauchi (Nagoya University, JP)

Joint work of Terauchi, Tachio; Yasuoka; Hirotoshi
Main reference H. Yasuoka, T. Terauchi, "On Bounding Problems of Quantitative Information Flow," Journal of Computer Security, 19(6):1029–1082, 2011, IOS Press.
URL http://dx.doi.org/10.3233/JCS-2011-0437

We present results on hardness of precisely checking and inferring a program's quantitative information flow (QIF). The results are presented from two perspectives:

1. verification theoretic view,
2. complexity theoretic view.

In 1.), we classify various QIF problems into program verification problem classes such as safety and liveness (and their recently-proposed extensions such as hypersafety and hyperliveness). This reveals that different QIF definitions, such as min entropy and Shannon entropy, often exhibit different hardness for some QIF problems. In 2.), we give complexity theoretic hardness results, focusing on boolean programs. The results uphold the classification given in 1.), and also show close connection of some of the QIF problems to counting problems.

3.26 Bayesian inference to evaluate information leakage in complex scenarios: case study mix-networks

Carmela Troncoso (Gradiant – Vigo, ES)

This work casts the trace analysis of anonymity systems, and in particular mix networks, in the context of Bayesian inference. A generative probabilistic model of mix network architectures is presented, that incorporates a number of attack techniques in the trace analysis literature. We use the model to build an Markov Chain Monte Carlo inference engine, that calculates the probabilities of who is talking to whom given an observation of network traces. We provide a thorough evaluation of its correctness and performance, and confirm that mix networks with realistic parameters are secure. This approach enables us to apply established information theoretic anonymity metrics on complex mix networks, and extract information from anonymised traces optimally. This work is further explained and evaluated in [1, 2].

References

1 C. Troncoso, G. Danezis, *The Bayesian Analysis of Mix Networks,* Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS 2009), E. Al-Shaer, S. Jha, A. D. Keromytis, Eds., ACM, pp. 369-379, 2009.

2 C. Troncoso, *Design and analysis methods for privacy technologies,* PhD thesis, Katholieke Universiteit Leuven. C. Diaz, B. Preneel (advisors), 189+17 pages, 2011.

3.27 Computer-Aided Proofs of Differential Privacy

Santiago Zanella Beguelin (Microsoft Research UK – Cambridge, GB)

Joint work of Zanella Beguelin, Santiago; Barthe, Gilles; Danezis, George; Grégoire, Benjamin; George; Kunz, César
URL http://easycrypt.gforge.inria.fr

Differential privacy permits the disclosure of noisy statistics of sensible data without compromising the privacy of individuals. Proving that a program that uses standard sanitization mechanisms guarantees differential privacy is relatively easy. However, proving that an arbitrary probabilistic program guarantees differential privacy is an error-prone task that calls for a principled approach and tool support. In this talk, I will show how the novel relational program logic of Barthe et al. [2] can be used to reason about differential privacy as well as its approximate and computational relaxations. Moreover, I will report on the implementation of a proof system for this logic into the EasyCrypt interactive prover [1], a tool that can be used to verify the security of cryptographic primitives in the computational model. The resulting integrated proof assistant allows users to verify privacy guarantees of general probabilistic programs that use cryptography, under computational assumptions. Finally, I will illustrate the use of this framework to verify that a two-party protocol for computing Hamming distance between bit-vectors yields two-sided privacy guarantees.

References

1 G. Barthe, B. Grégoire, S. Heraud, and S. Zanella Béguelin, "Computer-aided security proofs for the working cryptographer," in *Advances in Cryptology – CRYPTO 2011*, ser. Lecture Notes in Computer Science, vol. 6841. Heidelberg: Springer, 2011, pp. 71–90.

2 G. Barthe, B. Köpf, F. Olmedo, and S. Zanella Béguelin, "Probabilistic relational reasoning for differential privacy," in *39th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2012*. New York: ACM, 2012, pp. 97–110.

Participants

- Alessandro Aldini
Univ. of Urbino, IT
- Mario Alvim
University of Pennsylvania, US
- Anindya Banerjee
IMDEA Software Institute, ES
- Béatrice Bérard
UPMC, Lab. LIP6 – Paris, FR
- Arnar Birgisson
Chalmers UT – Göteborg, SE
- Michele Boreale
University of Firenze, IT
- Kostas Chatzikokolakis
Ecole Polytechnique – Palaiseau, FR
- Tom Chothia
University of Birmingham, GB
- David Clark
University College London, GB
- Jorge Cuellar
Siemens – München, DE
- Alessandra Di Pierro
Univ. degli Studi di Verona, IT
- Ehab ElSalamouny
Ecole Polytechnique – Palaiseau, FR
- Sardaouna Hamadou
Ecole Polytechnique – Palaiseau, FR
- Holger Hermanns
Universität des Saarlandes, DE
- Michael Hicks
University of Maryland – College Park, US
- Sebastian Hunt
City University – London, GB
- Daniel Kifer
Penn State University – University Park, US
- Boris Köpf
IMDEA Software Institute, ES
- Matteo Maffei
Universität des Saarlandes, DE
- Pasquale Malacaria
Queen Mary University of London, GB
- Fabio Martinelli
CNR – Pisa, IT
- Michael W. Mislove
Tulane University, US
- C. Carroll Morgan
UNSW – Sydney, AU
- John Mullins
Ecole Polytechnique – Montreal, CA
- Gethin Norman
University of Glasgow, GB
- Martin Ochoa
Siemens – München, DE
- Catuscia Palamidessi
Ecole Polytechnique – Palaiseau, FR
- Quoc Sang Phan
Queen Mary University of London, GB
- Alexander Pretschner
TU München, DE
- Andrey Rybalchenko
TU München, DE
- Mathieu Sassolas
Université Libre de Bruxelles, BE
- Vladimiro Sassone
University of Southampton, GB
- Fabrizio Smeraldi
Queen Mary University of London, GB
- Geoffrey Smith
Florida Int. Univ. – Miami, US
- Marco Stronati
Ecole Polytechnique – Palaiseau, FR
- Tachio Terauchi
Nagoya University, JP
- Carmela Troncoso
Gradiant – Vigo, ES
- Herbert Wiklicky
Imperial College London, GB
- Santiago Zanella Beguelin
Microsoft Research UK – Cambridge, GB