Information Security Inside Organizations

A Positive Model and Some Normative Arguments Based on New Institutional Economics

vorgelegt von Diplom-Informatiker
Frank Pallas

Von der Fakultät IV – Elektrotechnik und Informatik
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Ingenieurswissenschaften
– Dr. Ing. –

genehmigte Dissertation

Promotionsausschuss:

Prof. Dr. Jean-Pierre Seifert (Vorsitzender)
Prof. Dr. Bernd Lutterbeck (Berichter)
Prof. Dr. Louise Yngström (Berichterin)

Tag der wissenschaftlichen Aussprache: 09. Juli 2009

Berlin 2009
– D 83 –

Information Security Inside Organizations

A Positive Model and Some Normative Arguments Based on New Institutional Economics

Frank Pallas

Berlin, 2009

[T]he problems of IS/IT security management [are] tackled without drawing on existing theories, at least for the moment.

Fredrik Björck (2004)

Long-term messes become so familiar they look normal to most people living in them. In the absence of a means to fundamentally change the mess, people change themselves. They accept the mess as part of reality and build their worlds around it.

Peter J. Denning (2007)

Abstract

This work develops an abstract, theory-founded understanding of organization-internal information security. For this purpose, established knowledge from the field of information security is restructured on the basis of two different dimensions: The historical dimension distinguishes three "eras" of information security and relates them to concurrent changes of prevailing computing paradigms. The "security triangle" identifies and characterizes three different "meta-measures" for realizing information security inside organizations and highlights the existence of a higher-level regulatory framework.

Additionally, the work is based on principles from the field of New Institutional Economics. In particular, the concepts of information asymmetries, transaction costs and principal-agent relations are explicated as well as their relevance to the establishment of cooperation among individuals. Cooperation is in turn modeled as consisting of the two partial problems of coordination and motivation.

These theoretical foundations are then merged into an economically inspired positive model of information security inside organizations. The model provides abstract and theory-founded explanations for the changes of prevailing information security practices that happened in the past.

Besides this explanatory use, the positive model is also applied in a prospective manner. Current technological developments will presumably lead to increasingly "interwoven" computing structures and thus to another change of the prevailing computing paradigm. The application of the model to the changed givens suggests that now-established practices like behavioral guidelines or those means usually associated with the term "security culture" will prove inefficient and thus inadequate in the future. Organizations will therefore have to use alternative approaches or to modify existing ones for realizing information security under the changed circumstances.

Various possibilities for doing so have been suggested in the past. Some of these are evaluated on the basis of the economically inspired, positive model. This analysis leads to well-founded suggestions which of the approaches should be applied under what conditions. Furthermore, the economic understanding also supports the development of new approaches that have so far not been thought of. As a final aspect, the future role of the higher-level regulatory framework is illuminated. It is shown that this framework will have to be adopted to the upcoming changes in order to protect organizations from being forced to apply highly inefficient practices for compliance reasons alone.

Overall, the positive model developed in this work provides explanations for what can be observed in the field of organization-internal information security, allows for well-founded predictions about what can be expected for the future and leads to normative arguments regarding necessary changes of established approaches and practices. It might therefore prove valuable for future research in a multitude of ways.

Zusammenfassung

Die Arbeit entwickelt ein abstraktes, theoriebasiertes Verständnis organisationsinterner Informationssicherheit. Hierzu wird etabliertes Wissen aus dem Gebiet der Informationssicherheit auf Basis zweier unterschiedlicher Dimensionen restrukturiert: Die historische Dimension unterscheidet drei "Epochen" und setzt diese in Relation zu zeitgleichen Paradigmenwechseln der Computernutzung. Das "Sicherheitsdreieck" charakterisiert drei unterschiedliche "Meta-Instrumente" der Informationssicherheit und hebt die Existenz eines übergeordneten regulatorischen Rahmens hervor.

Zudem basiert die Arbeit auf Prinzipien der Neuen Institutionenökonomik. Insbesondere werden Informationsasymmetrien, Transaktionskosten und Prinzipal-Agenten-Beziehungen sowie deren Relevanz für die Realisierung von Kooperation betrachtet. Kooperation wird dabei als das Zusammenspiel von Koordination und Motivation modelliert.

Auf Basis dieser theoretischen Grundlagen wird daraufhin ein ökonomisch geprägtes, positives Modell der Informationssicherheit in Organisationen entwickelt. Das Modell liefert abstrakte und theoretisch fundierte Erklärungen für die in der Vergangenheit aufgetretenen Wechsel vorherrschender Vorgehensweisen der Informationssicherheit.

Neben diesem erklärenden Aspekt wird das Modell zudem prospektiv angewendet. Derzeitige technologische Entwicklungen werden voraussichtlich zu zunehmend "verflochtenen" informationstechnischen Strukturen und somit zu einem weiteren Paradigmenwechsel der Computernutzung führen. Die Anwendung des positiven Modells legt nahe, dass sich derzeit etablierte Vorgehensweisen wie Verhaltensrichtlinien oder üblicherweise mit dem Begriff "Sicherheitskultur" verbundene Maßnahmen zunehmend als ineffizient und damit unangemessen erweisen werden. Organisationen müssen daher entweder alternative Ansätze nutzen oder bestehende Vorgehensweisen anpassen.

Hierzu existieren bereits diverse Vorschläge, von denen einige auf Basis des ökonomisch geprägten, positiven Modells untersucht werden. Diese Analyse führt zu wohlfundierten Empfehlungen, welche Vorgehensweisen unter welchen Rahmenbedingungen angewendet werden sollten. Zudem unterstützt das ökonomische Verständnis die Entwicklung neuer, bislang noch nicht diskutierter Ansätze. Schlussendlich wird auch die zukünftige Rolle des regulatorischen Rahmens beleuchtet. Auch dieser bedarf der Anpassung an die zu erwartenden, veränderten Rahmenbedingungen um zu verhindern, dass Organisationen allein aus Compliance-Gründen zur Anwendung hochgradig ineffizienter Vorgehensweisen gezwungen werden.

Insgesamt liefert das entwickelte positive Modell damit Erklärungen für beobachtete Phänomene der organisationsinternen Informationssicherheit, ermöglicht wohlfundierte Vorhersagen zu erwartbaren zukünftigen Entwicklungen und führt zu normativen Aussagen zu notwendigen Änderungen etablierter Ansätze und Vorgehensweisen. Es kann sich damit für zukünftige Forschungen in vielerlei Hinsicht als nützlich erweisen.

Main Findings

This work examines the field of information security inside organizations from an abstract perspective. It is based on modern economic concepts like transaction costs, information asymmetries or agency theory. Starting from these these theoretical foundations, an economic reinterpretation of organization-internal information security is developed "from scratch". This reinterpretation then allows for a consistent, structured and theory-founded understanding of a multitude of security-related aspects.

Such fundamental and theory-based deliberations do, however, require several abstract concepts to be addressed, explained and consciously applied to the field of organization-internal information security. Unfortunately, this leads to a certain extensiveness having to be accepted for the sake of precise and well-founded analysis. But on the other hand, such extensive and highly abstract considerations might also form an obstacle for grasping the essential findings and implications that result from the abstract, economically inspired approach pursued herein.

Before steeping into the rather fundamental considerations, the main findings and implications of this work shall therefore be briefly explicated in the following:

Main Finding 1:
Information security inside organizations can be described, comprehended and analyzed on the basis of established concepts and theories from the field of modern economics.

As it will be shown throughout this work, well-known principles from the discipline of New Institutional Economics can be applied to the field of organization-internal information security in a multitude of ways. It is, for instance, possible to explain security-related behavior as being subject to the economic concept of "externalities" and to adopt established economic approaches for externality-related problems to the field of information security.

Based on such considerations, an understanding of organization-internal information security can be reached that is based on established theoretical foundations and that allows for a structured representation and perception of various security-related aspects. In particular, an economic understanding provides deeper insights into the interrelations between organizational science and information security than other approaches do.

Main Finding 2:
The realization of information security inside organizations is a problem of economic cooperation among individuals. Any such cooperation requires coordination and motivation.

Within such an economic understanding, all security-related efforts of an organization can be interpreted as pursuing the goal of cooperation among individual members. As the security-related behavior of any single member can affect the outcome of any other member and of the organization as a whole, organizations have to find ways for realizing cooperative – as compared to purely selfish – member behavior with regard to information security.

Such cooperation always requires two partial problems to be solved: coordination and motivation. Of these two, coordination refers to the identification of the "optimal" state of member behavior that would represent the highest overall value for the organization. This coordination outcome could, for instance, be that no member except salespersons should use public WLAN hotspots.

The second partial problem of motivation then refers to the enforcement of the coordination outcome. As individual members have various incentives to behave opportunistically instead of conforming to the identified "optimal" state, they have to be motivated to abstain from this opportunistic behavior and to behave in the collective interest of the organization instead. If, for instance, the coordination outcome states that non-salespersons should not use WLAN hotspots, this "optimal" state also has to be enforced to prevent the respective members from using a hotspot for selfish reasons alone.

In any case, coordination and motivation raise a certain amount of costs having to be borne by the organization. These costs are influenced by a multitude of factors that will be discussed in detail throughout this work. Without going more into detail, different modes of coordination lead to different cost structures and different motivational means do so, too. Besides realizing cooperation at all, an organization therefore also has to identify those approaches to coordination and motivation that are most suitable to its specific givens and requirements and that thereby provide the highest efficiency.

Main Finding 3:
An economic understanding of organization-internal information security allows to explain the changes in prevailing security practices that happened in the past.

The interpretation of security-related efforts as being aimed at cooperation among individual members provides an economic explanation for the changes of prevailing security paradigms that happened in the past. Different practices of coordination and motivation lead to different cost structures.

In the earliest days of isolated, large-scale computers, security-related cooperation among the different members of an organization only required to distinguish operators from non-operators and to enforce this distinction (or, more precisely, the respective behavior aspired for operators and non-operators) by means of physical protection.

The use of mainframe computers which were shared by diverse users at the same time then brought with it different changes that can be interpreted in an economic manner. In particular, the possibility of interrelations between the different members using such a shared system required coordination to be realized in a more detailed manner. This

fine-grained coordination process consequently resulted in fine-grained coordination outcomes that could not be enforced through the then-established physical means anymore. This need for enforcing more detailed coordination outcomes then led to the development of technical means of motivation like access control mechanisms.

The introduction of networked PCs, in turn, heightened the costs arising from the coordination process in a multitude of ways. Complexity was increased even further, coordination had to be realized under strong uncertainty and information asymmetries, and user capabilities were strongly broadened. Altogether, these developments would have led to cost increases of coordination that would possibly have outreached the value added by the use of PCs instead of mainframes.

The change of prevailing security practices that happened together with the switch from mainframes to PCs can be interpreted as a response to these changed economic conditions: By relying on more managerial approaches like behavioral guidelines or "security cultures", organizations could delegate ultimate decisions to the members themselves but could at the same time motivate these members to make decisions that reflect the collective interest of the organization. The different cost structures of such modes of cooperation might then have allowed organizations to actually profit from the use of PCs instead of mainframes at all.

An economic interpretation of organization-internal information security does thus at least provide a conclusive and theory-founded explanation for the past changes of prevailing security paradigms.

Main Finding 4:
Emerging technological developments will lead to significant changes of organizational structures. For the field of organization-internal information security, these changes will in turn constitute new challenges because of altered economic givens. The economic model that is developed in this work allows to comprehend these changes and challenges in an abstract and precise manner.

Ongoing developments toward higher member mobility, less static membership statuses, more outsourcing and the increased use of technologies like service-oriented architectures (SOA) or Software-as-a-Service (SaaS) blur the previously well-defined organizational boundaries in a multitude of ways. Instead of acting as isolated and stable entities, organizations will increasingly be interwoven with each other and consist of an ever-changing set of members and technical entities.

This does, of course, have implications for the necessary cooperation process and for the respective costs of coordination and motivation. With regard to coordination, complexity, uncertainties and information asymmetries will be increased even further as compared to the above-mentioned case of PC usage within well-defined and comparably static organizations. On the other hand, the strong asymmetries will, in a slightly different manner, also influence the costs of motivation. In the end, technological progress will thus lead to changed economic conditions and thereby imply new challenges for the realization of security-related cooperation.

Main Finding 5:
The prospective application of the economic model suggests that some established instruments of organization-internal information security will prove strongly inappropriate under the conditions having to be expected for the future. In particular, this is the case for practices usually subsumed under terms like "security culture" and "security guidelines".

The new economic preconditions implied by technological and organizational change do, first of all, imply that a highly detailed determination of aspired, security-related member behavior would prove strongly inefficient. As factors like complexity, uncertainty and information asymmetries will be even higher than it was the case for PC-based environments, "traditional" approaches based on a completely defined state of member behavior that is enforced through technical means would again seem inappropriate. At first sight, this might suggest organizations to use managerial approaches like behavioral guidelines and "security cultures" even more intensively than today.

However, as the economic properties of these practices would lead to significantly increased motivation costs under conditions of strong decentralization and interwovenness, behavioral guidelines and "security cultures" will presumably prove inefficient, too. Both approaches require a certain probability of noncooperative behavior to be actually detected. If this is not the case, individual members have no incentive to follow the respective rules – be they formal, law-like or rather informal, norms-like – and would be motivated to behave opportunistically instead.

In fact, such an increase of opportunistic behavior is exactly what can be expected to happen within strongly decentralized settings. As noncooperative behavior will become significantly less likely to be detected, the effectiveness of managerial instruments of motivation will be lowered and ultimately, the costs of motivation will presumably increase significantly. This, then, will again necessitate alternative practices of information security that better fit the newly arising conditions.

Main Finding 6:
The economic model can support the development and the prospective evaluation of alternative approaches to organization-internal information security. This leads to possible strategies that would otherwise presumably have been overlooked and that will – under the changed conditions that can be expected for the future – in all likelihood provide a higher efficiency than the now-established practices.

Due to the above-mentioned change of economic conditions under which security-related cooperation will presumably have to be realized in the future, alternative approaches need to be developed. These approaches will also have to solve the partial problems of coordination and motivation but will have to do so under significant information asymmetries existing between strongly decentralized individuals which are furthermore situated in ever-changing contexts.

A multitude of alternative approaches have already been proposed in the literature for overcoming the newly arising challenges: Context-enabled access restrictions,

overridable technical means of enforcement or even more sophisticated ideas based on "insecurity credits" can be mentioned in this respect. With regard to such already existing proposals, the economic understanding developed in this work can help assessing their viability through prospective evaluation and choosing the appropriate approach for a given organizational setting.

Furthermore, the economic perspective to information security and the upcoming challenges also makes suggestions for the development or refinement of further approaches. While, for instance, the approach of overridable technical means has so far only been suggested in combination with conscious ex-post inspections of an override's adequacy, one could also think of "tax-like" payments being levied for any such override. And finally, economic theory also provides first ideas for the development of future approaches to security-related cooperation that possibly go without some centralized instance and rather function in a more market-like manner. In the long run, such approaches will presumably be necessary because of hierarchical cooperation becoming highly inefficient in general. So far, however, such approaches to information security must be seen as first ideas that require extensive further consideration.

Main Finding 7:
Organizations are constrained in their choice of means for realizing information security by a higher-level regulatory framework. This framework would in its current form presumably avoid the application of novel approaches. Existing security-related regulations should therefore be reconsidered and adjusted to the changed givens having to be expected for the future. Organizations would otherwise be forced to consciously implement strongly inefficient procedures.

Whatever approaches were chosen or developed by an organization as response to the newly arising challenges, most of them would presumably conflict with the current regulatory framework in one way or another. This regulatory framework consist of a multitude of standards and legal regulations which prescribe some security-related practices and prohibit others. It thereby defines the "field of the game" that an organization must adhere to when choosing its security-related course of action.

The use of a novel approach would only make sense for an organization when older, now inappropriate practices can be abandoned at the same time. But doing so would in many cases not be compatible with the current regulatory framework. If, for example, the regulatory framework in the form of security standards obligates an organization to conduct awareness trainings, this organization might not substitute novel approaches for these awareness trainings but would rather maintain an outdated and largely inefficient practice for compliance reasons alone.

Besides this possible obligation to maintain outdated practices, the current regulatory framework might also prevent the application of new ones more directly. Some of the alternative approaches discussed in this work do, for instance, rest upon extensive logging of user activities and contextual information and might thereby raise conflicts with privacy-related legal regulations. In some cases, such restrictions might be perfectly reasonable because of considerations beyond efficiency. In other cases, however,

it might at least be questioned how much efficiency we are willing to give up in the name of other objectives.

Other alternative approaches which rest upon completely new strategies like explicit pricing of insecure behavior, for example, would simply "not fit" into the established regulatory framework. They would thereby presumably rise a multitude of further conflicts that could also prevent their adoption.

Generally speaking, the currently established regulatory framework prescribes currently established practices and could hamper the use of novel, more appropriate approaches in a multitude of ways. This does at least call for conscious consideration of possible adjustments. Again, the economic understanding of information security developed in this work could support such deliberations.

Besides these main findings, many further results are developed throughout the course of this work. The most important goal, however, is the establishment of a consistent, theory-founded, economic understanding of organization-internal information security. This understanding provides valuable and new insights and can serve as a comprehensive source of inspiration for thinking about organization-internal information security.

Acknowledgments

Like any other work of a comparable kind, this dissertation would not have been possible without the influence of a multitude of people. While there is no way to enumerate all of them, some contributors of particular importance shall be mentioned explicitly. As their individual influence is a variable that is hard to measure and to rate, they shall, in consistence with the citation order of authors practiced at our chair, be named alphabetically.

My first thanks thus go to Matthias Bärwolff – one of my colleagues and co-candidates – for giving me an initial explanation of the principal-agent-model that is so pivotal for this work, for answering countless questions while sitting right beside me and, most notably, for *not* having reined my economic ambitions.

Then I want to thank Kai Dietrich – a brilliant-minded student and colleague – for endless and extremely fruitful discussions and for taking the role of the devil's advocate again and again. If there were (what I actually do not hope for) only one person in the world grasping *everything* I want to say with this work, I think it would be Kai.

Timo Glaser – my second co-candidate, an inspiring entrepreneur and a great pragmatist – deserves appreciation for repeatedly bringing me back to earth when my thoughts unquestionably got overly abstract. Furthermore, Timo perseveringly read earlier drafts of this work and helped me improve it in a multitude of ways. And, last but not least, Timo repeatedly served as cake-supplier, thereby satisfying a priorly unknown demand and initiating spontaneous brainstorming sessions that always turned out to be highly constructive.

The support provided by Kei Ishii – another colleague who already crossed the doctoral line some years ago – cannot be put into written words. He made his invaluable contributions by simply being Kei Ishii. Those who know Kei will also know what I mean.

I am deeply grateful to Bernd Lutterbeck – my first doctoral advisor and the head of our research group – for offering me to leave a predestined career path as programmer of SCADA-visualization and to join his research group instead, for continuously encouraging me in my doing and for providing me with a working environment shaped by an uncommon level of independence, autonomy and intellectual liberty. With his unconventional way of thinking, Bernd made me recognize the importance of economics and motivated me to call established assumptions into question. Whenever his remarks seemed somewhat crazy to me, I realized their underlying wisdom days, weeks or even months later and was ultimately able to profit from them.

Oliver Raabe – a legal scholar who actually lives interdisciplinarity while others only talk about it – deserves gratitude for repeatedly serving as sparring partner in various discussions on rather fundamental issues of regulation and on the limits of the economic

approach. Willingly or unwillingly, Oliver made me rethink my argumentation in a multitude of ways.

I thank Karsten Weber – a wanderer between the worlds of computer science, philosophy and the law – for assuring me of being on the right path, for various discussions on scientific methodology and, not to forget, for his incredibly dry sense of humor.

And finally, I am indebted to Louise Yngström – a longtime proponent of interdisciplinary and theory-based information security research and my second advisor – for unconscious indirect and conscious direct inspiration and for her invaluable last minute support. Without Louise, I would presumably have been much less satisfied with the ultimate result of my doing.

Besides these professional supporters, I want to say "thank you" to all those people who accompanied me on my way over the past years and who always reminded me that there is a wonderful world beyond the university's boundaries. Especially during the often burdensome final months, Gela, Marcus and Torsten repeatedly persuaded me to come with them and go out into the wilderness for searching some geocaches or simply took me out for a beer. Even if I often wanted to work instead, I am now convinced that they were absolutely right. I would otherwise have gone insane.

But above all, I want to thank Beatrice, my wife, for being in cahoots with the above-mentioned other geocachers, for her endless patience and understanding over all the years and, in particular, for her consent when I got the offer to go on the scientific journey and to aspire intellectual fulfillment instead of money. Without her, I would not have been able to get through all that.

Contents

Introduction

This is a doctoral dissertation and its subject is, according to the title, information security. More precisely, this work is about information security *inside organizations* and discusses aspects that could, with significantly more prevalent terms but missing the actual point to a certain extent, also be interpreted as information security *management*.

Given these conditions, most readers would presumably expect me to begin my considerations with some generic remarks on the importance of information technology, its influence on nearly any aspect of our daily life and its unquestionable indispensability to the conduct of business activities. Furthermore, I would presumably be expected to make some comments on the relevance of security aspects for the ("effective and efficient") operation of organization-internal IT systems, to mention numerous statistics highlighting a "still unsatisfactory" level of information security to prevail in a multitude of organizations (and leading to "substantial losses" year after year) and, not to forget, to note that information security can nowadays not be seen as "solely technological issue" anymore.

I am, however, not going to do so. In fact, I already had such an introduction at hand but ultimately came to the conviction that it would hardly have been of any value for those readers already having opened and started to read this work. I will thus abstain from such common deliberations and use a different approach instead.

As I have been taught by someone who must know, legal scholars always start with a *case*. Only after having grasped what the case and its particular specifics are, they begin with the identification, analysis and explanation of relevant fundamentals and apply these to the given case. Within such an approach, the case serves various goals: It gives motivation to concern with more abstract subjects, provides evidence for the relevance of considerations being made and significantly heightens vividness. And last but not least, a case ensures a certain grounding of rather abstract reflections, thereby counteracting the risk of producing outcomes that hardly have any practical relevance. If there were only one thing that could be learned from legal scholars, I think it would be this approach to always base abstract considerations on a concrete case.

For these reasons, this work also begins with a case. Chapter 1 gives a short introduction to the use of public WLAN hotspots within a professional context and exposes a concrete problem arising from the need to realize security in this context. This problem is then used as starting point for our more abstract considerations and discussions. Furthermore, explicit remarks on the scope and the structure of this work and comments on the scientific approach being pursued will also be addressed throughout the course of this case-chapter.

Let us therefore embark on our journey and begin with our introductory case: The secure use of public WLAN hotspots.

Chapter 1

An Introductory Case: Public WLAN

Here then is the dot.

– Lawrence Lessig

Back in 2003, hardware vendor Intel released its "Centrino" platform, a set of hardware components for notebooks – consisting of a processor, a chipset and a WLAN[1] interface – that should make mobile computers faster and lighter and should provide longer battery life and more comfortable communications.

The launch was accompanied by an impressive marketing campaign. In public advertisements, Intel promoted its vision of how work would look like some years later: Employees would take their desks out to the roofs of office buildings or to golf courses and parks to conduct their work while enjoying sunny weather, fresh air and maybe a calming view over a pond. Mobile performance together with wireless connectivity would – as it was the vision – allow people to work when, where and how they want.

Generally speaking, the spots promoted a new kind of work, less bound to offices, more self-determined and – in the end – more enjoyable. The already existing ideas of mobile telework and "nomadic computing" (Kleinrock 1995, 1997; Lyytinen and Yoo 2002), which had formerly been practiced in smaller niches or in a less distinctive manner, suddenly got broad public attention and were widely considered as a possible alternative for a large number of employees.

Of course, the advertisements were consciously exaggerated. No one would have expected offices to become unnecessary. And of course, "Centrino" was not the first technology providing the promoted functionality off-the-shelf.[2] But nonetheless, the launch of "Centrino" can legitimately be seen as one of the key drivers for the increasing spread of WLAN and the more general shift from desktop to mobile computers that we observe today. Notebook sales have reached or even exceeded those of traditional desktop computers and hardly any notebook comes without WLAN capabilities. Ultimately, the use of these WLAN capabilities has – the advancing adoption of 2.5G

[1] WLAN (Wireless Local Area Network) is also known under the name of the IEEE standard it implements (IEEE 802.11) and as "WiFi", which stands for "Wireless Fidelity" and is primarily promoted by the industrial WiFi Alliance. Instead of the cumbersome name of the standard and the somewhat marketing-flavored "WiFi", the more neutral term "WLAN" will be used herein.

[2] Apple, for example, already offered optional WLAN-connectivity (which was called AirPort) for some models in 1999. Concerning energy-optimization, on the other hand, "Centrino" was in fact a significant step, even if long-running mobile computers existed before, too.

and 3G services notwithstanding – increasingly turned into an often indispensable business need. No hotel or convention center trying to attract professional customers can nowadays afford not to offer WLAN connectivity within their premises.

Hotels, convention centers and other commercial and privately organized players are, however, not the only source for obtaining publicly usable WLAN connectivity. In particular, there also is a multitude of initiatives for providing WLAN through a community-driven process and a lower number of public bodies providing WLAN connectivity as part of their basic infrastructure.[3] Altogether, these different models of provision lead to WLAN connectivity being broadly available for professional users.

Besides availability issues, a multitude of further aspects have to be considered with regard to the adoption of public WLAN hotspots within professional contexts. Topics like cost efficiency, bandwidth requirements, ease of use or the blurring distinction between work and leisure time could be mentioned in this respect. All these and many further issues could with good reasons be discussed in more detail.

The main topic of this work, however, is information security and in fact, security has been a key concern with regard to both – locally managed, organization-internal WLANs as well as externally provided hotspots – from the very beginnings of their use. These security issues were originally addressed by the WEP-protocol (Wired Equivalent Privacy) that allows to restrict access and to encrypt data traffic on the basis of pre-shared keys within locally managed networks.[4] But admittedly, after several flaws were found in the WEP protocol in 2001[5], it had to be considered insecure and was not viable for professional use anymore.[6] Alternative security mechanisms thus had to be developed for the secure use of WLAN in a professional context.

These alternative mechanisms have been developed over time and can today be regarded as being widely established. The prevailing approach, which is now used by a multitude of organizations to operate internal WLANs in a secure manner, shall thus be delineated in the following. Based on these considerations, we will subsequently proceed to the discussion of externally provided, public hotspots and the possibilities to use them securely. As we will see, there is a small but important difference between these two technical environments that has far-reaching implications for the use of public hotspots within professional and security-sensitive contexts.

[3] Even if these different models of ownership and provision unquestionably represent a highly interesting subject, they shall not be discussed explicitly herein.

[4] Furthermore, many products allowed and still allow to support authentication by means of MAC address filtering. MAC addresses identify network adapters and can thus be used to restrict network access to a certain set of registered devices. Nonetheless, MAC addresses can basically be forged or "spoofed" and MAC filtering can thus only be seen as secondary mechanism, supporting other primary methods of authentication.

[5] With the establishment of new WLAN encryption protocols such as WPA or WPA2, the shortcomings of WEP are nowadays usually seen as overcome. For the weaknesses of WEP, see especially Fluhrer, Mantin, and Shamir (2001) and Stubblefield, Ioannidis, and Rubin (2002). For further progress in decrypting WEP, see also Tews, Weinmann, and Pyshkin (2007)

[6] See, for example, Williams (2002) *"[D]eploying a wireless network [is like] putting an Ethernet jack in your parking lot and letting the world plug into your network."*

1.1 Security in Locally Managed WLANs

The nowadays established model for operating a locally managed WLAN has been described by Henry and Luo (2002, pp. 68 f) as well as by Williams (2002). In this model, WLAN access points that should extend a corporation's wired LAN are (logically) located in a DMZ[7] instead of directly connecting them to the internal network (see figure 1.1). This approach, however, necessitates additional techniques like those being established for traditional remote access – encrypted VPNs[8], for example – to be used for accessing the internal network via WLAN. On the other hand, using such remote access techniques also allows to *"let the incumbent security environment protect LAN assets"* within a local WLAN scenario (Williams 2002, p. 45) and thus make a unified procedure possible.[9] The same approach was also recommended by Housley and Arbaugh (2003, p. 34) and even by the Wi-Fi Alliance itself (Wi-Fi Alliance 2003b, p. 3).

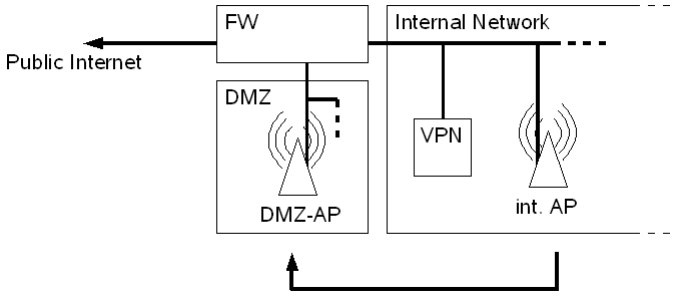


Figure 1.1: Location of access points in the DMZ instead of the internal network

To illustrate the functionality of this approach in more detail, the layered model of the internet protocol suite can be applied (see figure 1.2).[10] We have to consider

[7] DMZ stands for "DeMilitarized Zone" and refers to a *"network segment [...] located between protected and unprotected networks"* (Sheldon 2001, p. 522). DMZs are usually used to provide selected services to be accessed from unprotected public networks.

[8] The term "VPN" herein refers to what is termed a "*secure* VPN" by the VPN Consortium (2006): A network with traffic being encrypted and tunneled during transfer over public networks. A "*trusted* VPN" – as defined by the VPN Consortium – works without encryption.

[9] Nonetheless, making use of secondary mechanisms like MAC-filtering could still make sense to pevent externals from using the corporate WLAN, for example. See Williams (2002, p. 46): *"Putting the WLAN network outside of the firewall doesn't necessarily mean that itinerant passersby have free access to the Internet because you can still require authentication at Access Points."*

[10] The model of the internet protocol suite consists of four layers: The "applications" layer, the "transport" layer, the "network" layer and the "data link and physical" layer (See Sheldon 2001, pp. 654 f). The Internet Protocol (IP) is assigned to the network layer, TCP and UDP belong to the transport layer and protocols like HTTP, FTP, POP3, etc. represent the application layer, while the lowest layer represents numerous standards for basic connections like ethernet or even the different WLAN standards (802.11a, b, etc.). The 7-layer ISO OSI reference model (Zimmermann 1980) could also have been applied here, but this would have increased complexity without providing additional value.

five elements that are involved in the communication process: The user's device (UD, being equipped with WLAN capabilities), the WLAN access point (AP), the corporate firewall protecting the internal network from external attacks (FW), the VPN gateway (VPN) and the internal server providing the service the user wants to access – a groupware server, for example (SRV). To enable the user to access the internal server, a connection is established between the user's device and the WLAN access point (step a). Such connections take place at the lowest layer of the internet protocol suite (the data link and physical layer).[11] When this connection is established, the internal VPN server is approached. The firewall is configured to let all external traffic that is directed at the VPN server pass, the user gets authenticated by the VPN server[12] and the VPN connection is established (step b, connections being protected by encryption are highlighted by a gray rectangle). The protected VPN connection is assumed to be established on the network layer, like it is the case for IPSec, for example.[13] From now on, the user's device is *logically* located *inside* the internal network and can thus use any service being internally available via TCP or UDP connections (step c). The user can thus connect to the internal groupware server (application layer, step d) and work like being physically connected to the internal network.

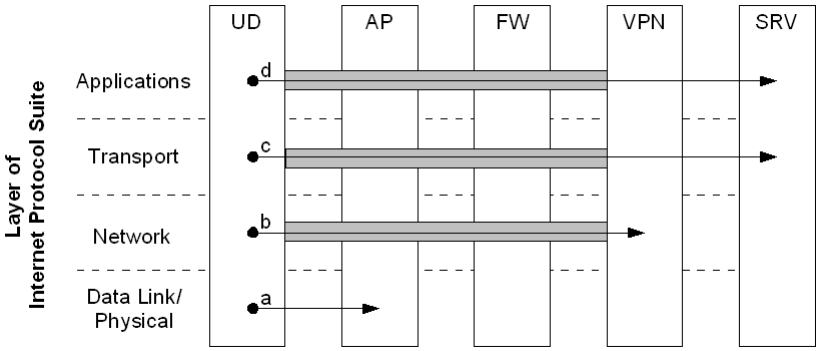


Figure 1.2: Secure access via local WLAN and VPN

Even if this approach is nowadays – with advanced and more secure successors of WEP like WPA or WPA2 being widely established – not indispensable anymore, it still allows for a unified procedure and infrastructure for accessing the internal network via a multitude of connections including local (DMZ-)WLAN as well as wired lines or 3G cellular services. Different from WPA or WPA2-based solutions, this generalized approach does not necessitate any further connection- or technology-specific precautions to be taken. For this reason, the use of VPN has nowadays established for nearly

[11] See Potter (2006, p. 52).

[12] For the general methodology, it does not matter by which means (a simple password, PKI mechanisms, etc.) the user gets authenticated.

[13] See Sheldon (2001, p. 683): *"IPSec [provides] security services at the IP layer in the Internet protocol stack."* Other VPN solutions differ marginally but without having an impact on the basic principle described here.

any kind of access from out of the local wired internal network.

Additionally, VPN solutions can be used to enforce the respective devices to be *always* logically placed in the *internal* network by routing any traffic above the physical and data-link layer through a VPN and thus through the internal network.[14] This approach of an "enforced VPN" thereby isolates the respective device from the network originally being connected to, circumvents any possibly insecure connection not being protected by the usual security infrastructure (the corporate firewall, for example) and thus drastically minimizes the risk of mobile devices being successfully attacked while residing in an "insecure" network.[15] For this reason, the methodology of an "enforced VPN" can be used to technically enforce network security policies like "no insecure connections above data link layer in and over unprotected networks", reliably isolating the respective devices from "insecure" networks and allowing to treat them like being internally connected from the network security perspective. "Enforced VPN" thus allows to widely expand the internal network to external connections (including WLAN) while still preserving the existing level of network security.

1.2 Security and Public WLAN Hotspots

For publicly usable WLAN, the use of VPN solutions is even more indispensable. To attract as many users as possible and to make access as simple as possible, hotspots *have to be* operated without making use of any of the existing security mechanisms like WEP, WPA or WPA2.[16] Any traffic between the user's device and the hotspot is thus interceptable by any nearby attacker as long as no additional precautions have been taken. Even service providers themselves thus *"recommend using a virtual private network [...] to keep traffic secure"* (Potter 2006, pp. 53 f). Furthermore, Godber and Dasgupta (2003) identified several additional threats[17] and consequently suggest *"to require all traffic to pass through a VPN to a trusted, secure, wired network"* (p. 430) instead of using a VPN only for *"'sensitive' or corporate traffic"* (p. 426) – a perfect equivalent of the above-mentioned policy of "no insecure connections above data link layer in and over unprotected networks" that would call for making use of an "enforced VPN".

[14] The respective VPN solutions typically either work by *automatically* establishing the VPN connection every time a network connection is available at the data-link layer and by redirecting all network traffic above the data-link layer through the VPN or they block any traffic above the respective layer as long as it is not routed through an established VPN connection.

[15] For the general threats possibly arising from devices that can be connected to "secure" as well as to "insecure" networks, see, for example the BSI (2004, T 3.41)

[16] All these mechanisms require some kind of pre-shared knowledge to be delivered between the operator and the would-be customers in order to allow users to connect to the hotspot. This delivery, together with the configuration efforts also being necessary, would put a burden on potential users and presumably keep many of them from using the hotspot at all. See also the Wi-Fi Alliance (2004a, p. 11): *"Security measures such as Wi-Fi Protected Access™ (WPA) and Wired Equivalent Privacy (WEP) require keys that are not easily distributed at public hotspots. In order to promote unhindered access and maximum use of their hotspots, venue owners rarely deploy these security measures."*

[17] The possibility of existing 'rogue' access points resulting from the network not authenticating to the accessing device, for example.

However, passing *all* traffic through a VPN or even using "enforced VPN" is impossible in most settings involving the access to public hotspots. Even if providers have a vital interest in operating their hotspots without activated encryption to attract as many customers as possible, they also have to restrict access to authenticated users – for billing purposes, for example. Operators thus have to maximize accessibility of their hotspots while at the same time having to realize a method for user authentication and access restriction. This dilemma is in most cases solved by implementing the *"Universal Access Method (UAM)"* (Wi-Fi Alliance 2004b, p. 5).

In this model, the hotspot has no access restrictions in place but is (logically) connected to the internet via an additional gateway. The gateway realizes what is called a *"captive portal"*:[18] In the first step, it completely blocks any network traffic except HTTP-requests. HTTP-requests are, instead of forwarding them to the respective website, answered with the login page of the hotspot operator, where the login credentials have to be entered.[19] Only after these login information have been verified, traffic blocking is turned off and access to the internet is granted.[20]

In terms of the internet protocol suite mentioned above, this architecture significantly changes the way a VPN connection can be established while using a public hotspot (see figure 1.3). Different from the model of a locally managed, organization-internal WLAN discussed above, we additionally have to consider the gateway (or the "captive portal") in this case (GAT). The first step is similar to a locally operated WLAN: The device connects to the hotspot at the physical/data-link layer (step a). After this connection is established, the gateway blocks any traffic until the user has authenticated. Instead of establishing the VPN connection, the user thus has to start an internet browser and make an HTTP request in order to be shown the respective login or advertisement page.[21] This operation takes place on the "applications" layer of the internet protocol suite (step b). Only after the respective page has been shown (and after potential login credentials have been transmitted), full internet access is granted, the VPN connection can be established (step c), and the internal server can be used (steps d and e).

With the "Universal Access Method" in place, it is thus impossible to have *all* traffic being passed through the protected VPN connection, as at least the "captive portal" page has to be retrieved *before* the VPN can be established at all. The successive steps for approaching the internal server thus do not strictly "move upwards" in the layer model one after another but rather "bounce" along the layer hierarchy. At first sight,

[18] See, e. g., Godber and Dasgupta (2002, p. 42), Hole, Dyrnes, and Thorsheim (2005, pp. 31 f) or Brunato and Severina (2005, p. 60).

[19] These credentials can be a username and password, a credit card number or even an access code from a scratch-off card.

[20] This method is not only used by commercial operators but also within municipal WLAN projects as well as in community-driven networks. As most of these networks rely on some kind of customership or membership, users have to be identified before being granted unrestricted access. And even in networks without membership, users are in most cases redirected to an advertisement-page before getting full internet access. Furthermore, legal aspects do in some cases play a certain role, too.

[21] See also Balachandran, Voelker, and Bahl (2003, p. 3): *"user authentication is done before procuring [...] a VPN connection"*.

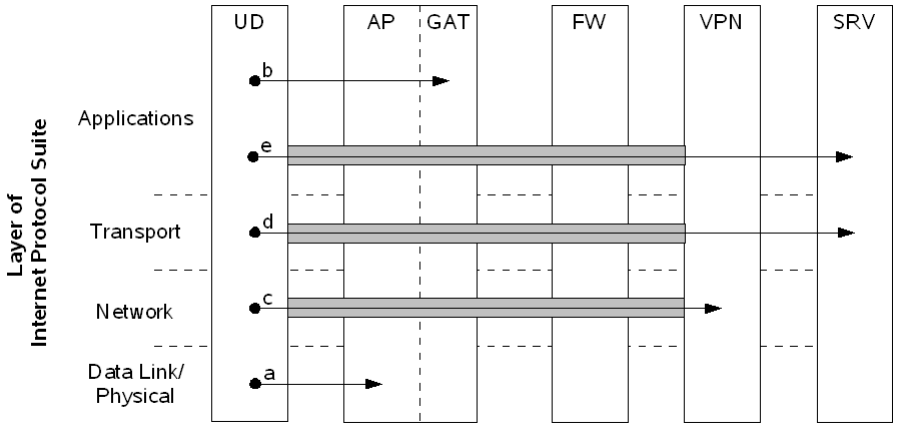


Figure 1.3: Secure access via restricted WLAN-hotspot and VPN

this might be a marginal difference, but the impact for practical network security is far-reaching.

From a general perspective, the described method involves the transfer of unfiltered – and thus untrusted – application level content (the login or advertisement page) onto the user device and its processing in the respective browser application. This content might contain malicious code and could thus generally pose a security risk for the device and for the internal network, accordingly.[22] Furthermore, VPN-usage cannot be enforced technically anymore and establishing the VPN has generally to be left to the user after he accessed the "captive portal" and possibly submitted the respective credentials.

This would, in turn, give the user the general possibility of connecting to and using "insecure" networks without establishing a VPN connection at all. For example, a user might abandon using a VPN because the corporate firewall blocks any access to certain internet sites he is interested in.[23] These websites could again contain malicious content and represent an additional threat to the security of the device and the whole internal network. Similar risks arise with regard to incoming connections[24] or in relation to the confidentiality of internal data.[25] All these and many more risks originally arise from the "Universal Access Method" requiring browser-based authen-

[22] This risk is even increased by the possibility of connecting to rogue access points.

[23] Another motivation for the user not to use the VPN connection could be found in performance aspects. See Hole et al. (2005, p. 31): *"To enable Web browsing, the traffic must first go through the VPN tunnel and the company intranet, before going back out on the Internet. This solution, however, might not be very efficient."*

[24] The device might only be equipped with a simple firewall in contrast to the enterprise firewall protecting the internal network.

[25] Think of accidental directory sharing or file uploads via webmail sites, for example. The mentioned risks might be mitigated by client side security precautions like virus scanners, local firewalls, etc. to a certain extent. But this does not eliminate the general risk of the typically more sophisticated security precautions in the internal network being circumvented.

tication, thereby making it impossible to pass *all* traffic through a VPN or to make use of an "enforced VPN". The "Universal Access Method" is thus basically incommensurate with a strict strategy of isolation between "secure" internal and "insecure" external networks.

Even if publicly usable hotspots are nowadays widely available, organizations could therefore still abstain from making use of them because of security objections. Nonetheless, there are several possibilities for solving this conflict.

1.3 Four Intuitive Approaches

Intuitively, one could identify four potential approaches for handling the described problem of secure access via public WLAN-hotspots in an organizational context:[26]

1. A general ban on WLAN-usage (or, at least, on the usage of WLAN-hotspots with restricted access and a required login) for the whole organization, possibly supported by technical measures like "enforced VPN" or hardware deactivation.

2. Technical modifications of hotspots and the according infrastructure, allowing for a lower-layer authentication of users and thus reestablishing the possibility of making "no insecure connections above the physical/data-link layer".

3. Technical solutions that automate the manual login process.

4. A more lax treatment of WLAN-usage, allowing short periods of unprotected connections but obligating users to manually establish protected connections as soon as possible in order to minimize the extent of "unprotected periods".

These approaches will be briefly discussed in the following. There might be good reasons to pursue any of them, but on the other hand every approach also imposes several side-effects that might – depending on the respective organization – result in considerable disadvantages.

1.3.1 General Ban

The easiest way to overcome the security problems arising from WLAN-usage is to abstain from accessing public hotspots at all. This "general ban" strategy typically arises from a strict policy of network isolation like the one mentioned above. It can be realized by an "acceptable use policy" which employees have to obey, by technical solutions like an "enforced VPN", or – and this is the most likely alternative – by both together.

The advantages of this approach are clear: The organization does not have to bear any additional security risks from WLAN-usage and has no additional expenditures – neither in the form of explicit payments nor in the form of additional precautions

[26] Of course, there might be further and especially more sophisticated approaches. Nonetheless, we will concentrate on these four to exemplify an organization's general options for dealing with conflicts like the one described herein.

having to be taken. The internal network can be protected by a strict strategy of isolation and network security can be demonstrably assured.

On the other hand, organizations with a "general ban" on the usage of public WLAN also have to abandon certain opportunities like an elimination of idle time or an increasing productivity and flexibility. They are not able to heighten service quality through better availability of means of communication. And finally, they might even suffer from lower employee satisfaction or from missed business opportunities accruing from unforeseen events.

The strategy of a "general ban" thus represents an extreme stance on network security, allowing organizations to achieve a well-defined course of action and to avoid additional risks and costs, while at the same time forcing them to give up the chance for miscellaneous possible benefits.

1.3.2 Change of Hotspot Infrastructure

As the best way of dealing with a problem or a conflict typically is its elimination, trying to eliminate the need for application-layer authentication seems promising. To allow for a process of establishing a VPN-connection which strictly "moves upwards" in the layer hierarchy, however, user authentication had to take place on the "physical/data-link layer" or on the "network layer" of the internet protocol suite.

In fact, such mechanisms are already arranged for in current IEEE standards. IEEE 802.1x-based authentication, for example, can also be used to restrict access to WLAN access points.[27] The authentication method defined by the standard takes place at the "data-link" layer of the internet protocol suite[28] and would thus perfectly meet the demand for authentication on one of the two lowest layers. As 802.1x-based authentication is typically used together with a RADIUS server, it could generally be employed for authenticating users across provider boundaries[29] and would thus allow for a well-structured way of user authentication without the drawbacks of the "Universal Access Method". Other comparable methods to realize authentication on one of the lower layers could also be thought of, but they would basically work in the same manner.

The benefits of such authentication methods are obvious. If such an authentication method would be available at the hotspot that should be accessed, establishing a VPN connection would be possible without having to transfer content over an unprotected higher-layer connection. Authentication could possibly be realized within the VPN application and the conflict between network security and hotspot usage would disappear.

[27] The 802.1x standard explicitly refers to WLAN networks several times and is intended to be used in this context. See especially IEEE (2001, p. 21): *"[...] in order to support the use of Port-based Network Access Control in IEEE 802.11 wireless LAN infrastructures."* With the IEEE WLAN standard 802.11i (IEEE 2004), 802.1x-based authentication was explicitly incorporated into the 802.11-family.

[28] For the 802.1x standard being located at the "data-link" layer, see especially the structural layer overview given by the IEEE (2001, p. iii).

[29] For WLAN roaming supported by RADIUS, see especially Wi-Fi Alliance (2003a).

Nonetheless, this scenario is based on the preconditions that pre-established accounts exist[30] and that hotspot providers actually support the alternative authentication method. To be usable at as many hotspots as possible, the different providers furthermore had to agree upon a standardized procedure and to allow users to roam between different providers with only one pre-established account. Even if such cooperations are already established among larger providers to a certain extent for the "Universal Access Method"[31], providers currently do not seem to be interested in establishing similar cooperations for lower-layer authentication.[32] In fact, there are not even signs that providers allow such authentication methods only inside their own networks.[33] In any case, the approach described herein would require larger changes in the existing hotspot infrastructure. These changes had to be made by the providers, which, in turn, do (at least currently) not seem to have an interest in doing so. Consequently, solving the conflict through wide establishment of lower-layer authentication has – at least for the moment – to be considered unrealistic.

Altogether, the strategy of changing the existing hotspot infrastructure by adding the possibility for lower-layer authentication seems promising and would allow for a secure use of public hotspots with strict network isolation. Nonetheless, this would necessitate strong cooperation between different providers to establish a standardized authentication model and to allow roaming. Even if such cooperation including roaming agreements already exists for the "Universal Access Method" and even if the according IEEE standards are well-established, providers do not seem to have an interest in offering such an alternative method of authentication. Lower-layer authentication would thus perfectly meet the demands for secure use of hotspots but has – at least

[30] See Matsunaga, Merino, Suzuki, and Katz (2003, p. 114): *"Although [802.1x] authentication methods work well in corporate WLAN environments, they exclude one-time credit-card authorization options and free advertisements in public WLANs, because they assume a pre-shared secret between user and network."*

[31] There is, for example, an established practice of roaming agreements between different WLAN providers, allowing customers of one provider to access the hotspots of another one with their already established account. For example, the Hotspot Directory's 2009 list of hotspots being usable via a BT Openzone account included 53.225 sites worldwide as of January 2009. Of these hotspots, only 2.973 were stated to be self-operated by BT Openzone. Other providers like iPass make extensive roaming agreements and in this way even reach considerable coverage without maintaining *any* hotspot on their own.

[32] In Germany, there was an initiative named "Greenspot" of the internet alliance eco to realize a clearing house for fostering cooperation of hotspot providers and for making 802.1x-based roaming possible. Nonetheless, the initiatives does not seem to be active anymore.

[33] One might speculate about the incentives especially of large providers not to offer such lower-layer authentication services. One reason might be that providers would not be able to show advertisements on the "login" pages anymore. This would especially be significant for advertisement-based offers. Other reasons might be found in the field of economies of scale, critical mass and collective action: Establishing an additional login infrastructure involves costs and might – in particular if it is based on a public key infrastructure – only pay off if enough other providers also establish a similar system. Finally, especially those providers having their origin in the telecommunications sector – and with T-Mobile, AT & T, Orange and many others, these are the most important ones – might even have an interest in lower-layer authentication *not* being available in order to give professional customers a reason to choose 3G data services, where authentication *is* realized on a lower layer and where strict network isolation is thus possible to establish. However, the real motivations not to offer lower-layer authentication remain unclear to a certain extent.

currently – to be considered unusable.

1.3.3 Login Automation

The approach of automating the login process represents – in addition to the concept of lower-layer authentication mentioned above – another alternative for authentication without the "Universal Access Method". Especially the largest providers often offer specialized and provider-specific client software that handles the credentials of a respective pre-established account and thereby allows logging on to the provider's hotspots without browser-based authentication.

Even if such specialized software would presumably also transfer authentication data on the application-layer, it would still minimize risks by not being vulnerable to any browser-based attacks, for example. Additionally, such specialized software could enhance usability as it allows for the integration of hotspot-logon and the establishment of the VPN-connection. In the case of such an integrated solution, the user would not have to care about submitting credentials before manually establishing a VPN tunnel but only had to push a button being labeled "establish a VPN connection over a hotspot", for example. And even if such an integration would not be realized, network isolation could be achieved to a certain extent by limiting higher-layer transfer over "unprotected" network connections to the respective logon application.[34]

However, for this mechanism to be usable not only for logging on to a single provider's hotspots, different providers would either have to use the same authentication mechanism – which would lead to the same issues as the idea of a standardized lower-layer authentication mentioned above – or any provider-specific mechanism had to be considered by the respective solution as well as a multitude of pre-established accounts and alternative identification/billing mechanisms. These requirements would make establishing a universally usable client software for hotspot authentication at least a complicated task and consequently, current client software for logging on to hotspots is typically provider-specific and thereby strongly limits connectivity options.[35]

Using a specific software for hotspot authentication could thus enhance network security (even if it would not necessarily eliminate higher-layer data transfer) but would at the same time strongly limit the number of usable hotspots as current solutions are typically provider-specific. To overcome these drawbacks, providers would have to agree upon a uniform authentication mechanism or any provider-specific mechanisms had to be considered separately. As both alternatives are currently unlikely to come true, logon-automation via a dedicated client software can only be used in conjunction

[34] Of course, one had additionally to trust the logon application not to perform any unwanted data transfer in this case.

[35] See, for example, Matsunaga et al. (2003, p. 155): *"Several WLAN providers deploy their own proprietary network access client [that uses] their own authentication protocols in the serving network"*. See also Evans, Wang, and Ewy (2006, p. 87): *"Typically these public access hotspots utilise a web-based front-end to an authentication system [...] or proprietary client software that reduces interoperability options."* Some limitations and challenges arising from such *"provider-specific modes of authentication"* were furthermore already mentioned by Balachandran et al. (2003, pp. 2 f).

with a significant limitation of the number of usable hotspots.

1.3.4 Lax Treatment

The approach of a more lax treatment would involve giving up any strict approach to unprotected communication on higher layers. Instead of prohibiting such connections or even making them impossible, an organization could basically accept higher layer connections but advice members to establish VPN network protection as soon as possible. For the case of using WLAN-hotspots, this would mean to allow the scenario of a browser-based authentication described above and to leave responsibility for the establishment of the VPN connection to the users.

This strategy would offer the highest flexibility of all alternatives described herein. Any publicly usable hotspot applying any current method of authentication could be utilized for establishing remote connections to the internal network. No restrictions had to be accepted in terms of a limited number of usable hotspots or as a result of a potential need for pre-established accounts. Users could handle even uncommon logon procedures being applied by single providers and no incompatibilities with any hotspot infrastructure had to be expected. The organization could thus entirely benefit from the wide availability of publicly usable hotspots to heighten productivity and service quality of mobile users.

On the other hand, the organization had to renounce the principle of strict network isolation and thus had to take the additional risks described in section 1.2: The possibility of malicious content being transferred to the device, the risk of such content being propagated into the internal network later and of course the possibility of (accidental as well as intentional) data leakages. The users' ability not to follow the instructions to establish a protected connection immediately after logging on to the hotspots and to generally use unprotected online connectivity instead heightens these risks even further.[36] Different from the other approaches, the organization's information security does in this case mainly depend on individual user behavior.

The strategy of a "lax treatment" thus offers the highest flexibility and the highest possibilities for profiting from mobile scenarios. It thereby represents the approach with the highest possible benefit. However, the approach also entails the highest potential risks for the organization's information security. And finally, the actual risk largely depends on individual user behavior.

1.3.5 Possible Approaches: Conclusion

Four possible approaches have been identified for confronting the fundamental conflict between usage of public hotspots and a general strategy of network isolation in an organizational context. One of these – the general change of the underlying technical authentication infrastructure of existing and future hotspots – eliminates the conflict

[36] Additionally, renouncing the principle of strict network isolation would presumably have strong impact to the organization's IT strategy on the whole in a multitude of ways. This aspect will be considered in later chapters.

through introduction of "lower layer authentication", but currently seems unrealistic. It can thus be excluded from further consideration.

The significant characteristics of the remaining three approaches can be related to each other as follows: The strictest approach of a "general ban" being technically enforced would perfectly hold up the principle of strong network isolation and would cause no additional burdens for the organization – neither in the form of additional risks having to be taken nor in the form of considerable additional measures having to be employed. On the other hand, the approach also prevents the organization from taking any advantage from the use of public hotspots with "UAM"-based access restriction.

The less strict approach of "automating the login procedure" through client applications breaches the principle of network isolation in the strict sense but keeps the resulting risks low as long as technical measures are used to restrict unprotected connections to the login-application alone. Additionally, using such login-applications will in practice always limit the number of usable hotspots to those being explicitly supported. The approach of login automation thus offers the general possibility of profiting from limited hotspot usage at all but in turn entails certain risks having to be taken by the organization.

And finally, the least strict approach of a "lax treatment" offers all possibilities of taking benefit from hotspot-usage but in turn results in the highest risks having to be taken and in users being responsible for minimizing the actual risk. Table 1.1 summarizes these relations.

Table 1.1: Strategies for using public WLAN – Security, opportunities and relevant entities

Strategy	Security	Opportunities	Relevant Entity
General Ban	high	none	technology
Login Automation	medium	medium	technology
Lax Treatment	low	high	users

1.4 Conclusion and Contribution to the "Abstract Puzzle"

Publicly usable WLAN nowadays features remarkable availability and has established as one of the preferred technologies for getting online connectivity and thus remote access to organization-internal networks while being on the move. However, it has also been shown that the prevailing method being used for necessary user-authentication at hotspots – the browser-based "Universal Access Method" – constitutes a fundamental conflict between usage of public WLAN and the approach of realizing security through strict network isolation and "enforced VPNs".

This conflict can be addressed by at least four strategies of which one is currently un-

realistic. The three remaining approaches all have benefits as well as drawbacks. The mentioned strategy of strict network isolation had to result in a "general ban" of public WLAN, which would minimize risks but would at the same time eliminate any opportunity of profiting from the use of public WLAN; The strategy of "login automation" via a dedicated software would break the principle of strict network isolation but still prevent most risks and would at the same time offer medium opportunities through at least some hotspots being usable; And finally, the approach of a "lax treatment" would open up the full opportunities of publicly usable WLAN while concurrently entailing the highest risks and making *users* – instead of technical solutions – responsible for security.

1.5 Toward the Abstract Puzzle

Based on these findings, the question is which approach should be chosen by a specific organization. For extreme cases, the answer is easy. An organization with vast security requirements and with hardly any value being derivable from the use of public WLAN would surely choose the strategy of a "general ban". An organization with hardly any damage possibly arising from potential security breaches but with high value that could be realized through using public WLAN would, in turn, presumably choose a more "lax treatment".

But such extreme relations of requirements are rather the exception than the rule. In most cases, an organization will have the requirement for "a certain" level of network security and would be able to derive "a certain" value from WLAN-usage. For such organizations, the answer to the question which strategy should be deployed will typically be a less definite one. In most cases, it could be summarized as "It depends."

The procedure typically being suggested for decision-making in such settings is a systematic cost-benefit analysis:[37] Any option entails a certain amount of "costs" on the one hand – i. e. the risk of notebooks getting infected with malware while using a public hotspot – and a certain amount of "benefits" – a higher productivity being derivable from the possible use of public hotspots, for example – on the other hand.[38] Abstractly speaking, these costs and benefits should be estimated as precise as possible and charged up against each other for any possible course of action. Based on these calculations, the option with the best cost-benefit ratio should then be chosen.[39]

[37] Schneier (2000, pp. 301 f), for example, suggests to calculate an *"annual loss expectancy (ALE)"* to determine whether a specific countermeasure should be taken or not and develops the formal method of building and calculating "attack trees" for this purpose (pp. 318 ff). Anderson (2008, p. 846) also mentions calculating the annual loss expectancy and other methods as being widely established *"to prioritize security expenditure [and to] provide a financial case for it to senior management."*

[38] The terms of "costs" and "benefits" – "direct" and "indirect" ones – are addressed more precisely and in more detail in section 4.2. Nonetheless, for understanding the general principle of a cost-benefit analysis being made in the area of information security, accepting the general coexistence of costs and benefits for any possible course of action is sufficient.

[39] In fact, such a cost-benefit analysis is also intuitively performed in the extreme cases mentioned above. With vast security requirements and hardly any value possibly resulting from an "insecure" course of action, it is typically clear without in-depth calculations that the "general ban" strategy

However, implementing a well-structured and objective cost-benefit analysis turns out to be unrealistic or at least problematic in many cases: Information security risks can in most cases not be determined exactly but only estimated.[40] The same is true for exactly quantifying the different beneficial opportunities resulting from a specific approach.

The situation gets even more complicated because of the different presented options being highly variable. Especially the option of a "lax treatment" could be combined with further countermeasures like formal regulations for employees not to use public WLAN without VPN-protection and substantial punishments for the case of non-compliance. Awareness-campaigns could also be used to sensitize employees and specific training could enable them to "behave securely". At least in theory, these measures could provide the approach of "lax treatment" with a substantial level of security while at the same time still offering most possibilities – and thus, most benefits – as the number of usable hotspots remains virtually unlimited.

And finally, as *people* are responsible for the actual level of security being present within the approach of a "lax treatment", the type of the organization might play an important role, too: In a small firm with a familial working atmosphere, users will presumably behave different than in a large corporation. These and a multitude of further factors make a systematic cost-benefit analysis of different options a highly complicated – if not impossible – task.

This is, of course, not only the case for the context of WLAN usage. The same considerations have to be made for the use of mobile storage devices, where carrying out sensitive data puts these data at risk but at the same time provides a certain benefit[41], for mobile working in general, where the derivable value for an organization has to be weighted up against threats like "shoulder-surfing"[42] or even for the employment of external consultants, which is done to derive a benefit from but which at the same time holds a certain risk of internal information being passed on to externals. In all these cases, an opportunity of gaining additional (business) value is opposed to additional information security risks having to be taken[43] and in all cases it is hardly possible to determine costs and benefits exactly. And finally, it has to be expected that for most organizations, a highly strict strategy *for* security and *against* additional opportunities would represent a suboptimal cost-benefit ratio.

---

offers the best cost-benefit ratio.

[40] See, for example, Anderson and Moore (2006, p. 610), identifying a fundamental *"difficulty in measuring information security risks"*. See also Anderson (2008, p. 847): *"[I]n real life, the process of producing [an exact annual loss expectancy (ALE)] table is all too often just iterative guesswork."*

[41] Again, there also exists the third option of device encryption, which provides a higher level of security but at the same time decreases functionality because of encrypted devices not being usable without the respective desktop application being installed on the computer the device should be connected to.

[42] "Shoulder surfing" refers to the risk of bystanders observing confidential information by looking onto a computer screen "over a user's shoulder", for example.

[43] This is not always the case for information security. There also exist measures – intrusion detection systems, for example – that heighten information security (and thus reduce risks) without significantly affecting opportunities. These cases are, as long as not mentioned explicitly, not further considered in this work.

1.5.1 The Need for Abstract Considerations

From an abstract point of view, the question having to be answered by organizations is thus not only how the possibility of using public WLAN should be treated but rather how possible benefits and information security risks should generally be balanced against each other in case of substantial uncertainty about costs and benefits and what kind of countermeasures should be used in which cases.

To give well-founded answers to such generalized questions, we need an abstract understanding of the different types of existing security countermeasures, their characteristics and their different ways of influencing information security inside an organization. We need ideas about which countermeasures are adequate for what kinds of requirements and which are not. And we need a better understanding of relations between different types of organizations and information security. In short: *We need a theoretic foundation for understanding and addressing the subject of information security inside organizations in a substantiated manner.*

One could assume that there should exist vast amounts of scientific literature being devoted to this issue. Of course, there is a great deal of literature that is dedicated to information security inside organizations in general, including technical guidelines, checklists, advice for applying security standards or simply codes of practice. This literature seems to be widely accepted amongst business professionals, but admittedly, it seldom uses an approach relying on *abstract* cognitions, concepts or principles.

What is still lacking is a well-established theoretical approach that provides abstract models, generalized problem types and structured methods for handling information security inside organizations. Up to now, we do only scarcely understand the underlying abstract principles of our doing. We cannot explain in a substantiated manner which strategy is promising under certain circumstances and which one is not. We have only few theoretically founded ideas about interrelations between information security practices and organizational structures. And, last but definitely not least, we have no well-established, *teachable* concept for understanding, explaining and realizing information security inside organizations that goes beyond swotting existing standards or frameworks. Such theoretical approaches have definitely played a minor role up to now.[44] From a scholarly perspective, this is at least unsatisfactory.

1.5.2 Scientific Approach

For understanding and addressing information security inside organizations in an abstract manner, we can thus scarcely rely on existing and well-established theoretical

[44] Björck (2004, p. 3), for example, identified only 25 out of almost 5000 papers from academic security journals containing the word "theory" in title, abstract or keywords. Hong, Chi, Chao, and Tang (2003, p. 243) mention a *"lack of information security management theory"* and Cavusoglu, Cavusoglu, Son, and Benbasat (2005) state that *"scholarly literature has been all but silent on developing [...] theoretical frameworks for organizational security management."* Some of the few exceptions from this general lack include the approach based on systems-theory and cybernetics that is represented by authors like Yngström (1996) or Kowalski (1994) and the idea of thinking about information security on the basis of institutional theory (Björck 2004). None of these does, however, seem to have gained widespread adoption so far.

foundations but rather have to build our own. Large parts of this work will therefore attend to the development of such an abstract, theory-founded understanding of information security inside organizations.

In view of the often long-winded and seldom fruitful nature of remarks on scientific methodology, only some brief notes on the approach pursued in this respect shall be given.[45] First of all, this work is for most parts based on *economic theory*. Starting from the fundamental assumption that economic theory and economic principles can profitably be applied to a multitude of behavioral aspects – including those that seem to be purely non-economic at first sight (Becker 1978) – we analyze the field of information security inside organizations from an economic point of view. Generally speaking, such an economic approach to information security can nowadays be regarded as well-established. The yearly workshop on the economics of information security, for instance, was initiated in 2002 by such outstanding personalities as Ross Anderson (2002), Bruce Schneier (2002) and Hal Varian (2004) and is now one of the most-renowned international conferences on information security.[46]

This fact notwithstanding, there is up to now no established and consolidated framework that explains organization-internal information security in an appropriate, consistent and teachable manner from an economic perspective. At least to a certain extent, the development of such a framework shall therefore be brought forward herein. For this purpose, we do in particular seize the ideas and basic assumptions of *new institutional economics*, namely the existence of transaction costs (Coase 1937, 1960; Williamson 1975, 1985), information asymmetries (Akerlof 1970) and the specific conditions existing within principal-agent-relationships (Jensen and Meckling 1976). As it is widely accepted in the field of economics, these concepts and theories, which are highly interwoven with each other, form the very foundations of economic interaction among individual players and therefore simply *must* be taken into account for the development of a viable economic model of information security inside organizations. As we will see throughout this work, their application to the field of organization-internal information security does, together with an economic understanding of human behavior[47] and of organizational structures[48], in fact lead us to insights and explanations that would otherwise possibly have gone unnoticed.

Doing so requires these concepts to be initially considered in a certain extensiveness to ensure a *substantiated* economic treatment of organization-internal information security throughout the later chapters. Our findings and results will, however, justify these somewhat tedious theoretical efforts. Different from various other approaches to organization-internal information security, we will be able to make our findings on the basis of an established and well-established theoretical fundament and we will be

[45] In so doing, we try to confine ourself to the most important aspects and to abstain from elementary remarks that would primarily be of historical interest. Those readers missing such fundamental comments might refer to any established reference explicitly being dedicated to scientific methodology itself.

[46] See, for example, `http://weis09.infosecon.net` [16.03.2009].

[47] The abstract and partially economic foundations of law and norms which will be incorporated in later chapters (Ellickson 1991; Posner 1997, etc.) can also be subsumed under this category.

[48] See figure 1.4.

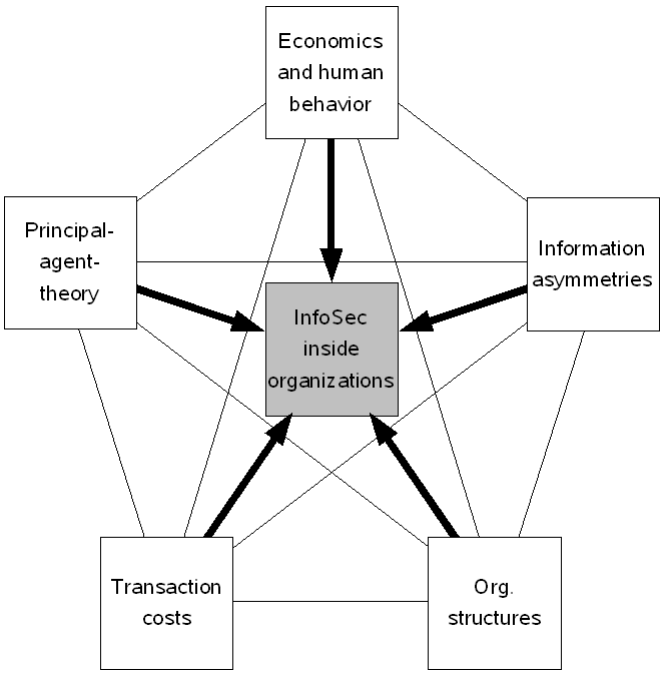


Figure 1.4: Theoretical foundations from the field of economics.

able to develop results that would at first sight seem somewhat counterintuitive.

Within such an economically inspired thinking, we do, secondly, follow the methodological approach of *positive economics* as depicted by, for instance, Friedman (1953): Consistently with our aim of developing an abstract and theory-founded understanding of information security inside organizations, positive economics is primarily concerned with the development of abstract models or theories describing and explaining *"what is"*.[49] Besides describing and explaining this "what is", any such model or theory also serves as a *"language"* or, metaphorically speaking, *"as a filing system for organizing empirical material and facilitating our understanding of it"* (Friedman 1953, p. 183). During the course of this work, we will develop a model of exactly this kind to establish an intellectual basis for considering organization-internal information security in an abstract, theory-based way.

Even if this already represents a reasonable end in itself – think of the problem of teaching information security in a consistent and understandable manner – such positive models or theories are usually developed in order to allow for well-founded predictions to be made for the case of changed circumstances.[50] The quality of these predictions being made on the basis of a certain positive model or theory is, in turn, the *only* valid criterion for evaluating the quality of the respective model or theory itself.[51]

[49] See Friedman (1953, p. 181), citing John Neville Keynes.
[50] See Friedman (1953, pp. 181, 205).
[51] See Friedman (1953, p. 184): *"[A positive] theory is to be judged by its predictive power for the class*

Mapped to the field of information security inside organizations, the positive model to be developed thus does not only have to describe and explain the current status quo of information security practice (the *"what is"*) but must also allow us to make valid predictions about expectable implications for the case of relevant circumstances being changed.

Finally, any such positive model or theory can also be used to derive *normative* arguments: Such statements (*"what ought to be"*) are necessarily based on *"prediction[s] about the consequences of doing one thing rather than another"* which, in turn, *"must be based [...] on positive economics"* (Friedman 1953, p. 181). Any normative statement should therefore rest upon well-founded positive models and predictions.

For the field of information security inside organizations, this implies that any normative argument on some course of action having to be taken or some practice having to be adopted (the *"what ought to be"*) should be based on conscious and explicit positive deliberations which, in turn, should result from some well-founded positive model or theory. This can – due to the above-mentioned broad absence of theoretical approaches in general – not be assumed as given today. On the other hand, however, the mentioned relation between positive models and normative arguments also implies that developing a positive model of "information security inside organizations" will presumably allow us to make better normative decisions than we would be able to without having such a positive model at hand.

Given these methodological considerations, we come to the following structure for the remainder of this work.

1.5.3 Structure

The first step toward the development of an economically inspired, positive model of information security inside organizations lies in a conceptualizing framework for grasping the current status quo within this field. This framework is developed in chapter 2 on the basis of historic considerations as well as on a restructured view onto the characteristics of established security mechanisms and practices. The outcome of these reflections are two representations of different aspects of organization-internal information security which our positive model will be based upon. Furthermore, we need to understand the basic concepts of organizations themselves in order to be able to make well-founded statements about information security *inside organizations*. These basic concepts are presented in chapter 3. In particular, this is where the various above-mentioned concepts from new institutional economics are introduced. These fundamental economic considerations then constitute the second pillar upon which we will later built our positive model. And finally, as costs and benefits – which play an important role for information security – are basically economic terms, we need some insights into selected, already established economic aspects of information security. These are addressed in chapter 4 and will serve various goals throughout our subsequent considerations. Together, these three chapters constitute the first part of this work which establishes the theoretical and abstract foundations of our

of phenomena it is intended to 'explain.'"

economically inspired, positive model of information security inside organizations (see figure 1.5).

Based on these fundamentals, we will then develop our positive model of information security inside organizations in part II. In particular, we identify organization-internal information security as typical cooperation problem and demonstrate that this problem is currently approached by hierarchical means (chapter 5). We then take up the historical representation of information security practices, combine it with the economic concept of hierarchical coordination costs and merge these two into an abstract understanding of security-related *coordination* in chapter 6. Chapter 7 does the same for the characteristics of information security mechanisms and hierarchical motivation costs to derive an abstract representation of security-related *motivation*. Together, these two make up our positive model of information security inside organizations, which will then for demonstrative purposes be applied to the initial case of using public WLAN hotspots (chapter 8).

Part III will then re-generalize from this specific case to broader developments that can be expected for the future. As suggested by the methodological remarks above, our positive model will be used to make predictions about the consequences of these developments for the field of organization-internal information security (chapter 9). Chapter 10 is then devoted to the normative implications of these predictions. Based on our positive model, we will discuss different possible approaches for tackling the imminent challenges for information security and estimate their appropriateness. Furthermore, we will develop some normative arguments regarding the rather "macroscopic" regulatory framework of organization-internal information security in this chapter. Chapter 11 summarizes our results, contains some notes on the limitations of our findings and ultimately motivates for further work that might be built upon our arguments.

Figure 1.5 gives an overview of this structure. This figure will be repeated at the beginning of every single chapter to always allow for an immediate positioning within the abstract logical structure, for a direct grasp of the status reached so far and for an easy estimation of the considerations that will follow.

As we will see throughout all the following chapters, the field of organization-internal information security can very well be addressed in a theory-based manner. As long as we stay open-minded enough to accept contributions from other disciplines to be worthwhile, we will repeatedly see things in a slightly different light than before. And if everything goes well, we might probably even be able to teach organizational information security in a different, more understandable and more scientific way.

May this humble contribution fall on fertile grounds and influence our view to information security inside organizations – in one way or another.

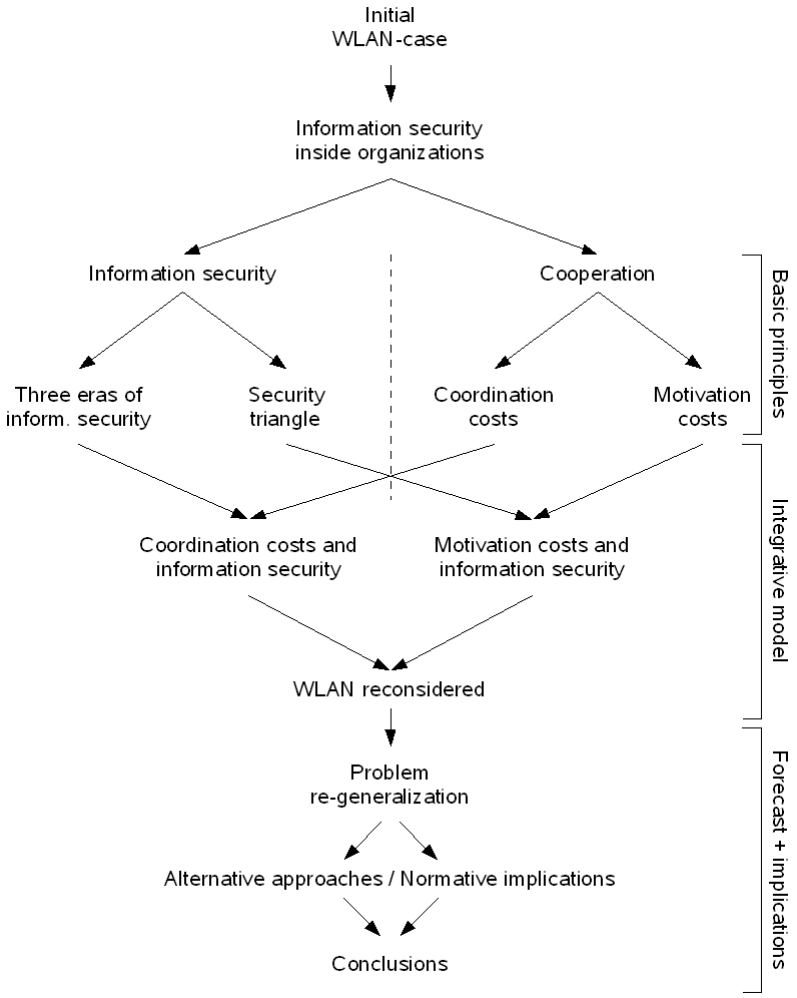


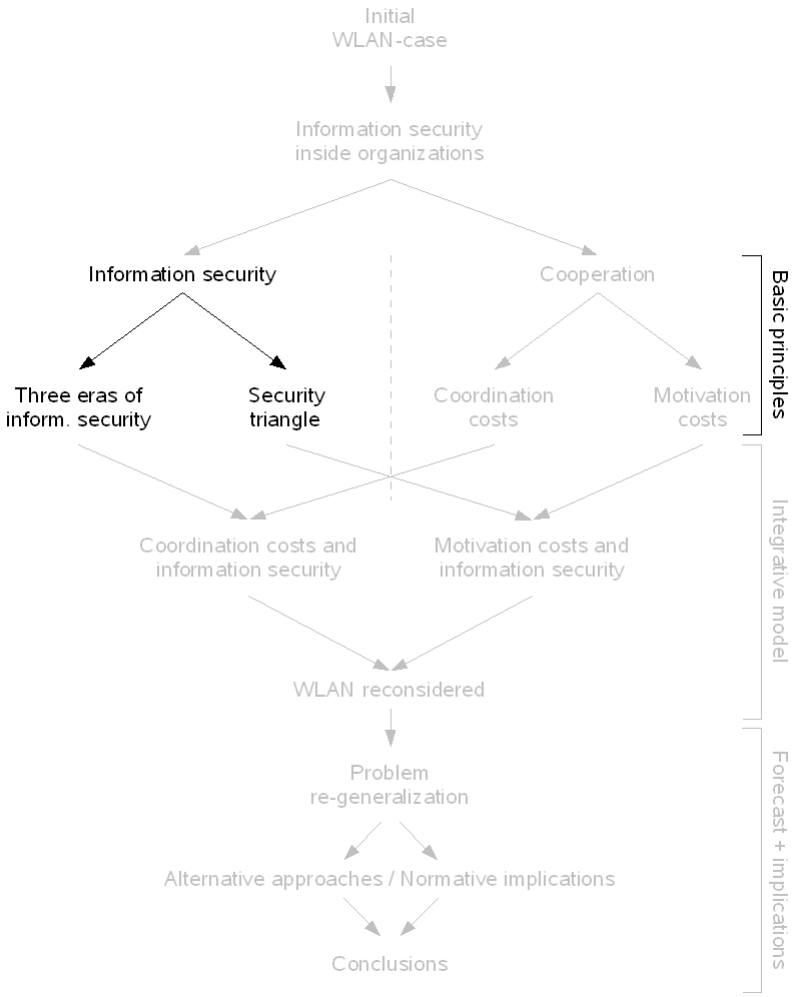
Figure 1.5: Logical structure of the work.

Part I

Information Security, Economics, and the Nature of Organizations: The Basic Principles

Chapter 2

Information Security in Organizations: Status Quo



Chapter 2

Information Security in Organizations: Status Quo

First make your model as simple as possible, then generalize it.

– Hal Varian

As outlined above above, this chapter is dedicated to the development of an abstract framework for discussing information security inside organizations in a theoretical manner. This is done by distinguishing and generalizing existing approaches on the basis of some significant characteristics. The idea is to identify different *types* or *classes* of established measures for information security and to discover common attributes that can be assigned to the different classes.

In a first step, this is done on the basis of an existing historical model from which an alternative model is derived (section 2.1). As, however, chronology is not the only possible way for identifying and distinguishing different aspects of information security, an additional model is developed in section 2.2 that distinguishes between different *"meta-measures"* for information security and identifies their different specifics. These abstract representations will then be used and expanded in later chapters to develop our positive model of information security inside organizations and to support the discussion of possible approaches for current and future challenges of information security.

2.1 The Historical Dimension - Waves and Eras

When trying to identify higher-level patterns in the field of information security, taking the historical point of view proves highly valuable as it allows to consider established approaches in conjunction with the respective context that led to their establishment instead of simply taking them for granted. Considering the historic development of challenges and approaches is thus our first step for analyzing the status quo of information security inside organizations on an abstract level.

Such a historical perspective was taken by von Solms (2000) to shape his model of "Three Waves of Information Security".[1] The main proposition of this model is

[1] There also exist other models for distinguishing different approaches to information security on a

that information security developed over time in three distinct waves, each one representing another dominating approach for information security. These three waves are characterized by von Solms as follows:

1. **The technical wave:** In this first wave, information security was solely shaped by technical approaches. The main idea was that technical means can solve all potential problems arising in the field of information security.

2. **The management wave:** Besides technical means, organizational structures like security managers were introduced and established during this wave. Also, this wave led to more involvement and attention of the top-management and furthermore resulted in organizations introducing formal regularities (like security policies) for information security.

3. **The institutionalizing wave:** Based on the two former waves, standardized procedures like best practices and the use of and certification against generally accepted standards became more important during the emergence of this wave. The introduction of information security metrics and other approaches should allow for a structured, repeatable and – to a certain extend – deterministic approach to information security. Also, employees were recognized as being relevant for information security, which led to the postulation to establish *"a culture of information security"* (von Solms 2000, p. 618).

Within this model, the waves and the measures being used by organizations are not said to replace but rather to complement each other. Even if the "Second Wave" stands for the growing importance being imputed to the management of security, this would, for instance, not be sustainable without still making use of technical measures and even enhancing them. Each emergence of a new wave thus represents the realization that hitherto existing mechanisms did not *suffice* to fulfill newly arising needs and that *additional* mechanisms are needed. Based on this realization, methodically different approaches got developed that better suited the respective new kinds of requirements (see figure 2.1 for a graphical representation of this relationship).

Through this principle, more and more *types of problems* could be addressed through making use of more and more different *types of measures*. The measures developed in the different "waves" therefore are, in economic terms, not strategic substitutes but rather strategic complements. In the following, these three waves will be examined and critically discussed in more detail.[2]

Additionally, several aspects that were not considered by von Solms are also taken into account to allow for deeper insights into the respective changes. Primarily, these additional considerations refer to the changes in the computing paradigms prevailing

historical basis. Siponen (2006b), for instance, distinguishes the three generations of *"premethodology, methodology, and post-methodology"*. However, the model of von Solms is – as we will see – much more expedient for the scope of this work.

[2] In doing so, aspects of information security that do not refer to the use of information *technology* are explicitly ignored. Short overviews of information security from the times before the computer are for example given by Russell and Gangemi (1991, p. 24) or Singh (1999).

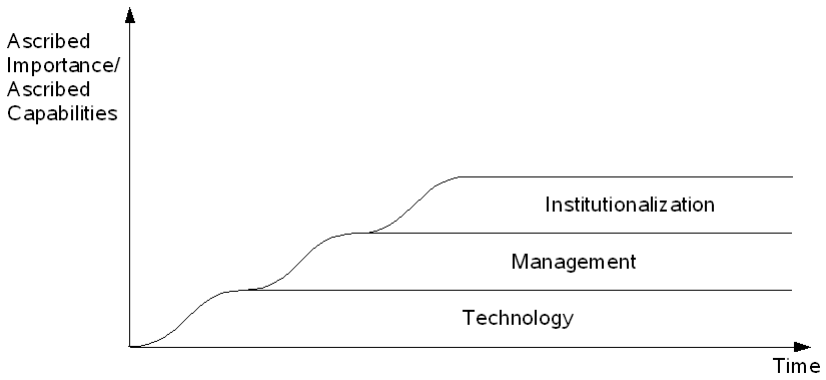


Figure 2.1: "Three Waves of Information Security" - Graphical representation based on von Solms (2000)

during the different "Waves". Based on the discussion and the additional considerations, we will then develop an alternative historical model which uses the term of "Eras" instead of "Waves" to be distinguishable from von Solms' model.

2.1.1 The First Wave of Information Security - Technology

The "First Wave" of information security represents the purely technical approach to information security that is, according to von Solms (2000, p. 615), rooted in the era of *"mainframe based"* computing. Earlier uses of information technology are not considered by von Solms. These were strongly shaped by isolated systems which had only been used in very few areas and for very specific purposes. They were not interconnected with each other and were only operated by highly specialized personnel. In this early era of "isolated computing", information security was – if at all – exclusively realized by means of what would today be called "physical protection": Thick walls, heavy locks and human guards watching over computers.[3]

But by the time, computers prove useful for many different purposes and providing their capabilities to more users became increasingly important. Making use of new concepts and technologies like time-sharing and networking, the introduction of mainframe systems made this possible: Those systems were typically used by many different users simultaneously (instead of the former model of batch-processing) but still condensed all the organization's capabilities for storing, accessing and modifying information at one central instance. The fact of multiple users accessing the same computing resources simultaneously led to the identification of – at that time – new security problems that could not be solved by the established means of physical protection.[4] A report for the US Department of Defense (Ware 1970) can be seen as

[3] See, for example, Ware (1970, p. 1): *"A basic principle underlying the security of computer systems has traditionally been that of isolation—simply removing the entire system to a physical environment in which penetrability is acceptably minimized."*

[4] Consequently, the use of such *"resource-sharing computer systems"* was at that times effectively

representative for discussing this new kind of problems and thus as reflecting the way of approaching information security that is subsumed under von Solms' "First Wave". In this report, five groups of "leakage points" and according protective measures are identified (Ware 1970, pp. 4 ff):

- **Physical surroundings** were still seen as being relevant for information security. Areas containing computing equipment should be protected according to the highest security level of data being handled on the system.[5] This physical protection should not only be applied to the mainframe itself but also to connected periphery like terminals, printers, etc.

- **Hardware leakage points** had also been known from the era of "isolated computing" and had by that time mainly been addressed by shielding the location of the isolated system against tapping. With the emergence of mainframes that could be used through remote terminals, shielding got more difficult and more expensive. The then-established mechanisms were thus not sufficient to satisfy new needs arising from a new technology. *Technical* protection of remote hardware was the favored approach to encounter these newly arising class of problems.

- **Software leakage points** refer to issues that arise from improper protection against interference between different users and processes due to the complexity of operating systems and supplementing programs. The most-concerned threats were undetected modifications of software and access rights resulting from *"incomplete design"* (Ware 1970, p. 8). Facing such threats was, in turn, mainly recognized as an issue of *"proper design"* (id.) and thus of *technical* means, too.

- **Communication leakage points** partially represent a subset of the hardware leakage points already mentioned above. As it is the case for remote terminals themselves, communication channels between these terminals and the central mainframe also introduced new and formerly unknown security threats due to the decreasing possibilities of physical protection. Like for the protection of terminals, *technical* means were favored for protecting communication channels, too.

- **Organizational leakage points** were also recognized during this early era of information security, but were primarily associated with the process of assigning a clearance level to those people who were granted access to the systems. Other organizational means like those gaining importance during the later "waves" were not considered at all or only to a very small extent.

prohibited at least in some military environments (Ware 1970, p. vi).

[5] Those measures of physical protection were already known from the earlier era of "isolated computing". But as mentioned above, new security approaches do not substitute existing ones but rather extend them.

Resulting from these issues and concepts dominating information security discussions, the developed approaches were in most cases strictly technical.[6] In this field, large efforts were accomplished during the respective times. Groundbreaking models and mechanisms like the Bell-LaPadula access model[7] or the concept of public-key cryptography[8] were developed and are still used today.

This is what von Solms identifies as the "First Wave": The main goals during the era of mainframe computing were reliable *technical* solutions for *"isolating any given individual from all elements of the system to which he has no need for access"* (Ware 1970, p. 8) and dependable *technical or physical* ways to protect the newly introduced peripheral components of the mainframe-computing environment (esp. remote terminals and communication lines) to allow for the treatment of confidential data on such systems. Other aspects of Information security were also recognized, but were at the same time assigned significantly less importance.[9]

Overall, the main point of criticism that can be assigned to von Solms' version of the "First Wave" is the non-consideration of mechanisms being used for information security before the mainframe-era. During these times, isolated systems were used and information security was mainly achieved through physical means. This approach has to be distinguished from the technical approach that shapes von Solms' "First Wave" and our alternative model will therefore include a "First Era" of isolated systems, that was – in terms of information security – strongly shaped by physical protection.

The facts shaping von Solms' "First Wave" – the use of mainframe systems and the technical approaches of protection – can be conveyed to a "Second Era" of our model without modification. The reason for such new approaches to information security can be found in a major shift of the prevailing computing paradigm. For mainframe-systems being used through remote-terminals, the established means of physical protection did not suffice and additional technical means emerged as a response to the shift from isolated to mainframe-systems.

[6] See also Saltzer (1974), discussing *"access control lists, hierarchical control of access specifications, identification and authentication of users, and primary memory protection"* as the key mechanisms used to protect data in the Multics operating system, which represents the "state of the art" of that times.

[7] For a concluding overview, see Bell and La Padula (1976).

[8] See, for example, Diffie and Hellman (1976).

[9] For example, Ware (1970) spends nineteen pages for "Technical Recommendations" alone (pp. 26-45) but only eleven pages for "Policy Considerations and Recommendations" (14-25) and 2 pages for "Management and Administrative Control" (46-47). Additionally, large portions of the two "nontechnical" parts refer to procedural aspects of the technical recommendations.

Conclusion:
The first "Era" of information security was shaped by the use of *isolated systems* and the according means of *physical* protection. The second "Era" can be characterized by the introduction of *mainframe systems* which could be used by multiple users concurrently via remote terminals. The resulting main intentions for information security were a *strong separation* of different users and the protection of the newly introduced *peripheral components*, which could not be protected by the established model of isolation anymore. Most of the measures being introduced during this "Second Era" originated from the domain of *technical* protection.

2.1.2 The Second Wave of Information Security - Management

Von Solms' "Second Wave" of information security stands for the growing importance of managerial approaches for information security. He himself mentions *"Information security policies, information security managers and organizational structures for information security"* as shaping this "Second Wave" and identifies *"[t]he development of distributed computing, and later the Internet, WWW and E-commerce"* as driving forces which *"catapulted information security onto boardroom tables"* (von Solms 2000, p. 615).

These statements, however, merely explain why information security got the *managers' attention*: The growing importance of information technology on the whole and therewith the growing importance of information security unquestionably was the reason for the managements recognition of and interest in information security. But this does not provide an explanation for the emergence of the mentioned *different approaches* to information security during this wave. As it was the case with the shift from physical to technical approaches, a possible explanation could be derived from another major shift of the prevailing computing paradigm.

2.1.2.1 Historic Considerations

The appearance of managerial approaches is identified by von Solms as taking place *"from about early eighties"* (von Solms 2000, p. 615), which would be simultaneous to another major shift in computer usage: the growing spread of minicomputers or, as they got called later, PCs within organizations. Even if these computers were originally thought of as a technology that could enhance mainframe-based computing[10], support *"process control, manufacturing and laboratories"* (Henzel 1971, p. 7) or act as *"security controllers"* for larger mainframe computers (Hoffman 1977, pp. 108 f), they were also used as a replacement for the established remote terminals. This change was mainly shaped by the emergence of business applications for Minicomputers during the late 1970s and early 1980s.[11]

[10] See, for example, Ball (1971), stating that minicomputers could be used to optimize bandwidth usage of communication lines or to convert character codes between mainframes and remote terminals.

[11] Usually, the introduction of the VisiCalc spreadsheet-application in 1979 is recognized as the initial event for business usage of Minicomputers like the Apple][or, later, the IBM PC.

Being equipped with own computing power, storing capabilities and I/O-functionality, PCs moved data processing away from centralized facilities onto the desks of the users.[12] Computers were no more only operated by multiple users from their desks remotely but were rather placed and used locally on the desks – including a significant move of capabilities from central instances to decentralized entities. On the one hand, this move included computing power itself, but on the other hand, capabilities for (short- and long-term) data storage as well as those for extracting or exporting data from the closed systems (by using printers or floppy disks, for example) were relocated to the desks, too. All these capabilities had formerly been operated within the central, well-controlled and well-protected mainframe environment. Businesses did of course profit from this new freedom and flexibility – they wouldn't have used it otherwise. With PCs being used instead of centralized mainframes, more members were able to make use of computing capabilities than before and costs were reduced significantly.[13]

In their earliest applications, these PCs were used in a strictly isolated manner. Neither did they possess any networking capabilities nor were they used by different members simultaneously. In a way, the earliest PCs thus represented yet another incarnation of the isolated systems from the first era – even if these "isolated systems" were now accessed by "ordinary" members. Just like the isolated systems from the first era, PCs were therefore primarily considered as having to be protected by physical means. Even if there seems to have been some initial confusion about how to actually protect the newly arrived personal computers[14], PCs were soon considered as usual office equipment that could be adequately protected by physical means like locked office doors or removable storage media being locked away in specialized cabinets.[15]

The real challenges, however, resulted from the introduction of network-enabled PCs, combining local storage and processing capabilities with remote access to centralized databases or file repositories. This newly arising paradigm of "distributed computing" brought to light the boundaries of the established physical and technical approaches to information security. On the one hand, basic physical protection alone would not have sufficed anymore because of the possibility of connecting to the central instance and accessing the data stored there, but on the other hand, the established

---

For the significant role of VisiCalc, see, for example, Benson (1983, pp. 39 f). See also `http://www.danbricklin.com/history/intro.htm` [15.02.2009] and links included therein.

[12] See, for example, Shepherd (1977, p. 70), distinguishing the former distribution of *"access"* to centralized *"logic and memory"* from the upcoming distribution of logic and memory themselves.

[13] See, for example, Carr (2005, p. 69): *"[The PC] dispersed the power of computing to individuals, spurred ingenuity [and] increased personal productivity."* For a simple costs comparison between Mainframe-based and PC-based reporting and planning, see Benson (1983, p. 40). For a more theoretical discussion on the advantages of "end-user computing" based on economic agency theory, see also Gurbaxani and Kemerer (1990).

[14] Murray (1986, p. 2), for example, had to explicitly argue against treating PCs in *exactly* the same manner as the well-established large "computers": *"[Personal computers have to be protected] the way we protect copying machines. [...] [S]ometimes that surprises people because their first identification of the personal computer is not as a piece of equipment but rather as a computer. And everybody knows how you protect a computers right? You place computers in highly specialized environments. [...]"*

[15] See, for example, Murray (1984, p. 299): *"The simplest way to protect confidential data on a fixed file is to lock the room in which the file is kept."*

technical measures for *"isolating any given individual from all elements of the system to which he has no need for access"* (Ware 1970, p. 8) could also not be easily transferred.

There are different reasons for the non-transferability of the rigorous scheme of technical isolation, ranging from simple technical problems[16] to rather social or psychological reasons.[17] But presumably most important was that introducing the complete set of technical countermeasures known from the use of mainframe systems would have resulted in a significant loss of freedom and flexibility which users were already accustomed to. To realize the *full* potential of PCs, users simply *had* a "need for access" to all or at least most of the "elements of the system". And if the main goal provided by technical measures was access restriction, they thus could not be applied to PCs without loosing at least parts of the benefits provided by them through freedom and flexibility.[18] Even if the transfer of technical countermeasures would basically have been possible, it would thus have resulted in considerable drawbacks by taking away key benefits of the newly establishing paradigm of networked, distributed computing.[19]

Again, progress in computing technology gave new opportunities to organizations but at the same time introduced new information security issues which could not be addressed properly through the existing approaches. And again, this did not mean that the existing approaches – in this case the physical and the technical ones – lost their right to exist due to the emergence of the new technology of PCs. They still existed and were strongly needed. But PCs changed the way computers were used and therewith led to information security problems that could not be solved by the utilization of physical and technical means *alone*. Thus, a new approach was needed to meet the new challenges.

2.1.2.2 Discussion

Von Solms' "Second Wave" – which was stated to have fallen into the same time with the increasing use of minicomputers or PCs – is mainly characterized by the introduction of three new instruments for enhancing information security: *"Information security policies, information security managers and organizational structures for information security"* (von Solms 2000, p. 615). The latter two could be interpreted as usual effects of structuring in a field that gains more relevance. The growing importance of information technology and thus of information security within organizations in general could have caused them, too. The pure existence of information security managers and structures alone does not represent a methodically new approach to information security and thus would hardly legitimate the identification of a new "Era".

[16] For instance, the prevailing operating systems were basically designed to allow complete access and technical systems allowed users to change the used operating system by themselves. Even there existed ideas to prevent user initiated changes of the operating system, these did not establish. See, for example, Murray (1984, p. 299).

[17] See, for example Benson (1983, p. 43), referring to users simply liking *"the total control they have over their own computer environment."*

[18] See also Holden (1986, p. 28), postulating not to *"try to take away in the name of security what technology has made available"*.

[19] We will examine this aspect and the underlying abstract concepts in more detail later in chapter 5.2 and especially in section 6.3.

This is not the case for information security policies. The term of a "security policy" itself is ambiguous to a certain extent. According to Anderson, Stajano, and Lee (2001, p. 3), an information security policy can be defined as *"a set of high-level documents that state* precisely *what goals the protection mechanisms are to achieve"* (emph. added). Based on this definition, the authors proceed with an understanding of security policies as documents that are formulated in order to derive technical solutions from. By doing so, they treat policies as some kind of a *"specification"* (Anderson et al. 2001, p. 37) from which a protection profile and, from there on, technical implementation details can be deduced.[20] From such a point of view, a "security policy" is primarily a part of the requirements analysis for the implementation of technical means for information security. In this case, "information security policies" would only represent an improvement to the established model of "security by technology" and thus give little reason to identify a new paradigm or "Wave" of information security.

On the other hand, "information security policies" can also be seen as *"[d]eterrents"*, which *"clarify what constitutes legitimate use of the information system"* (Wiant 2005). According to Andress (2004, p. 5), security policies *"define how a company approaches security, how employees should handle security, and how certain situations will be addressed."* Such information security policies include much more than technical issues and also address aspects of human behavior and define a code of conduct for the use of computers inside an organization.

As one can see, there are many different ways of understanding the term "information security policy". The probably most concise understanding, which will also be referred to herein, was given by Karyda, Kiountouzis, and Kokolakis (2005): They distinguish *"security policies"* and *"security guidelines"*, where policies cover a much wider area and are partially *"'translated' in guidelines"*. If understood that way, security policies are not only the basis for the design of technical measures but also the basis for security guidelines that influence human behavior. These guidelines, in turn, actually represent a methodically new approach to information security which can be used to identify the emergence of a new "Era" of information security.

2.1.2.3 Implications

The existing physical and technical means did not suffice to address the new aspects of information security arising from the increased use of PCs inside organizations. With the introduction of information security guidelines derived from security policies, the physical and technical approaches got enhanced by the introduction of *written* (instead of technically implemented) rules which instructed the users how to use their computers to ensure security. These rules could be interpreted as some kind of "intra-organizational computer usage laws"[21] and thereby clearly differ from the at that time established approaches of physical and technical security.

Thus, even if von Solms himself also mentions security managers and organizational

[20] Nonetheless, Anderson et al. (2001, p. 37) also state that *"[e]specially at the highest levels the policy functions as a means of communication"*.

[21] A deeper discussion of similarities between information security policies and judicial concepts of laws will be given in section 7.1.

structures as characterizing his "Second Wave", it can be mainly characterized by the *introduction of nontechnical, more regulatory approaches* to information security as a response to the increasing use of minicomputers and PCs. This addition of a qualitatively new type of measures as well as the shift in the prevailing computing paradigm also allows the identification of an additional "Era" of distributed systems for our alternative model.

Conclusion:
The shift to the third "Era" of information security took place as response to the increasing use of *Minicomputers and PCs*. These gave freedom and flexibility to the users and resulted in benefits for the organizations. The resulting main challenge for organizational information security was to keep these benefits while still providing adequate information security. Therefore, established information security approaches got expanded by *nontechnical, more regulatory means* like information security guidelines.

2.1.3 The Third Wave of Information Security - Institutionalization

The "Third Wave" was, according to von Solms, shaped by four main aspects (von Solms 2000, p. 616): information security standardization and certification, the creation of an information security culture and the implementation of information security metrics.

Of these, "standardization" refers to the development of internationally accepted standards for information security. Von Solms explicitly mentions ISO / IEC 17799[22] as an example. This standard includes sections regarding all of the aforementioned significant classes of measures: physical security, technical security and the security policy. Such standards and other best-practice collections, which are also subsumed under the label of "standardization" by von Solms, got developed to *"provide a sort of baseline for information security"* (von Solms 2000, p. 616) and not to *"precisely and 100% indicate"* (id.) how a certain organization should act in the field of information security. Even with existing standards, Information security still has to consider specific requirements and conditions. But the introduction of standards and best-practices for information security provided the persons in charge with the possibility to check for not having missed any highly important aspect.

The concept of "certification" is strongly coupled with the concept of standardization. Like an internal information security officer could check the existing measures against an internationally accepted standard, external parties can do so, too. With this process of auditing, organizations can receive an official certificate and thereby affirm to have established a certain level of information security to other organizations or individuals. This certification can be voluntary – for example, to get higher customer confidence than a competitor[23] – or even obligatory – for being allowed to

[22] The Standard has been derived from the British standard BS 7799-1 and is now also known as ISO 27002.

[23] See also von Solms (2000, p. 616): *"Companies wanted to know how good their information security was, how they compared to other companies."*

handle certain kinds of data, for example.[24] In any case, certifications allow externals, to a certain extent, to estimate the degree of information security being established inside an organization.

The estimation of the current level of information security being in place inside an organization was also the objective for the introduction of internal information security metrics. Overall, the mechanisms being used within the concept of information security metrics are the same that are also used for the certification mentioned above.[25] The main differences are that reviews are not conducted on a *"periodic basis"* (von Solms 2000, p. 618) but rather continuously and that reviews are not executed by people but are tried to be implemented technically.

2.1.3.1 Historic Considerations

These three aspects mentioned by von Solms as shaping the "Third Wave" of information security can be called into question. Even if von Solms (2000, p. 616) states that standardization, certification and security metrics emerged due to *"top management continuously asking about progress and results"*, at least standardization and certification existed long before the time von Solms places it.[26] In fact, they had already been present during the mainframe era and were of high importance then, too. But as the view had been a physical or a technical one at that times, standards and certifications also took only physical or technical aspects into consideration.

For instance, the US National TEMPEST Standard, which refers to electromagnetic radiations and therefore regards physical means, was, according to Russell and Gangemi (1991, p. 37), already published in 1970.[27] Also, an industrial TEMPEST program was set up in 1974 to *"[o]utline criteria for testing equipment"* and to *"[c]ertify vendor equipment that successfully meets the TEMPEST standard"* (id.). Thus, a standard and a certification process already existed for the area of *physical* protection in 1974.

Another example, which refers to technical means, is the Digital Encryption Standard (DES). It was, according to Russell and Gangemi (1991, p. 36), *"adopted as a Federal Information Processing Standard [...] in 1977"*. Like the TEMPEST standard, DES did also not refer to organizational aspects, but with DES, the *technical* instrument of encryption was standardized. To be bought by US government agencies, any commercial product had to use this encryption standard. And government agencies in turn had no choice of buying other products than those using DES.

And finally, the TCSEC standard, which is widely known as "the Orange Book", also existed long before the "Third Wave". It was originally published in 1983 and repub-

[24] Additionally, one could also mention the "quasi-obligatory" case, where an organization is forced to get certificated to get a chance on the market at all.

[25] See von Solms (2000, p. 620): *"The BS 7799 Code of Practice mentioned above is used as the reference framework against which the measurements are made."*

[26] As the "Third Wave" was stated to be *"presently picking up momentum"* (von Solms 2000, p. 616) in 2000, it could be guessed to have started in the mid-nineties.

[27] The original document was labeled "National Communications Security Informations Memorandum 5000" (NACSIM 5000) and had for long time been classified. A declassified version from 1982 is available at `http://cryptome.info/0001/nacsim-5000.htm` [15.02.2009]

lished as TCSEC in 1985 by the US Department of Defense. Its use was mandatory for all technical security evaluation activities inside the DoD.[28]

2.1.3.2 Discussion

Those standards mentioned above should suffice to demonstrate that standardization and certification are no invention of the time that is associated to the "Third Wave" by von Solms. In fact, standards also existed during earlier times, even if they always concentrated on the type of instruments being considered relevant in the according era. The only new aspect of the "Third Wave" in this regard is that standards and certifications also became prominent for those mechanisms that evolved during the "Second Wave"[29] – like they became prominent for the mechanisms of physical and technical protection throughout the times associated to the "Second Wave".

The use of security metrics, in turn, has not been widely known during earlier times. Nonetheless, as mentioned above, the underlying concept is of no groundbreaking innovation that could be compared to the shift from physical to technical measures or the shift from technical to managerial approaches to information security. Basically, security metrics rest on the concept of certification and use them in a slightly different way. "Certifications" are conducted continuously, not periodically, and they are conducted by technical artifacts and not by human auditors.

Finally, the "Third Wave" is labeled the wave of "Institutionalization" by von Solms. From an abstract view, the term "institutionalization" could be understood as the process of evolvement and strengthening of institutions, whereas the term "institutions" itself has been defined as *"the humanly devised constraints that shape human interaction"* (North 1990, p. 3).Thus, if understood this way, an "Institutionalizational Wave" in information security had to refer to the evolvement and strengthening of "constraints", shaping human interaction with regard to information security. The emergence of such constraints are obviously not limited to the "Wave" considered here. Physical means like thick walls as well as technical means like access restrictions had always constrained human choice and thus shaped human interaction. It does not seem like the economic understanding of "institutions" had been on the authors mind.

As used by von Solms (2000, p. 616), the term "institutionalization" refers to a process which leads to *"information security [becoming] a natural aspect of the day to day activities of all employees of the company"*. Together with the examples as given by the author and as discussed above, it seems like the "Third Wave" should – instead

[28] See Department of Defense (1985, p. 2): *"This publication [...] is mandatory for use by all DoD Components in carrying out [automatic data processing] system technical security evaluation activities applicable to the processing and storage of classified and other sensitive DoD information and applications"*. See also Russell and Gangemi (1991, p. 35).

[29] For instance, the OECD *"Guidelines for the Security of Information Systems"* of 1992 (OECD 1992) recognized a need for *"administrative, organisational, [...] and legal"* approaches to information security. All of these could be ascribed to the non-technical, rather managerial approaches that emerged during the "Second Wave" (see above). The OECD guidelines of 1992 might therefore stand for one of the earliest standardization processes with regard to "Second Wave"-practices. Note, however, that the OECD guidelines already mentioned principles like *"awareness"* or *"accountability"* which would rather be ascribed to the "Third Wave", too.

of the potentially misleading term of "Institutionalization" – better be understood as a "Wave of Consolidation", where aspects already known from former times got consolidated, standardized and widely established. As this process – the evolvement of standards and certifications against these standards – had also taken place earlier, and as the usage of security metrics mainly represents a continuous internal certification process, the identification of a distinct wave on the basis of "standardization", "certification" and "metrics" seems questionable. These concepts do not prove to be innovative enough to mark a significant shift in thinking about information security inside organizations.

What, then, remains from the "Third Wave"? Von Solms also mentions, as last determining factor for the identification of a new "Wave", the cultivation of *"an Information Security Culture Right Throughout a Company"*. He reasons this by *"[t]he realization that employees are in most cases the biggest danger to a company's IT systems"* and states that this *"human dimension of information security cannot totally be solved by technical and procedural measures"* (von Solms 2000, p. 618).

The growing importance of this *"human dimension"* can again be reasoned by the ongoing changes of computer usage. Even if there was no revolutionary paradigm shift like the ones identified above, the trend of decentralization and of users obtaining more possibilities continued. Newly introduced operating systems[30] – like it had been the case with the introduction of minicomputers in general – increased the users' freedom and flexibility even more. Additionally, the Internet got important for many users' daily work.

Due to these changes, it was not possible anymore to define completely which activities are allowed and which ones are not – neither technically nor through written policies or guidelines. As the users got more possibilities and more power, they also got more responsibility for information security. To make their users aware of this responsibility, organizations started to conduct information security awareness programs and campaigns to teach and to sensitize their users.

These activities had two main intentions: First, the users should have an own interest in information security. They should be aware that the wealth of their organization, and thus their own wealth, strongly depends on information security. And second, the users should be *enabled to behave securely* by teaching them existing risks and the respective ways to avoid or to counteract them. Together, this should result in what von Solms (2000, p. 618) calls an *"Information Security Culture"*.

2.1.3.3 Implications

Again, this human-centric approach to information security strongly differs from those being favored during the earlier eras of isolated and mainframe-based computing. But the difference to the managerial approaches from the "Second Wave" is less significant: Both concentrate on *people* instead of technology itself. And both do not try to make insecure behavior *impossible* but rather try to prevent insecure behavior through

[30] For example, Microsoft Windows NT 4, at that time revolutionary from the professional user's view, was released in 1996.

other means – by law-like instructions or by fostering a security-aware organizational culture.[31]

Both, instructions as well as organizational culture, are broadly established as usual management tools. This would argue for subsuming both aspects under one "Managerial Era" instead of assigning them to different "Waves" like von Solms does.[32] From a historic point of view, written guidelines and the idea of an "information security culture" can, as the previous discussion showed, both be linked to the growing decentralization and thus to the rising freedom, flexibility and responsibility of computer users themselves.

Altogether, the "Third Wave" of information security as proposed by von Solms (2000) can be mainly characterized by the realization that human aspects play an important role for information security. This led to increasing attempts to promote an information security culture which should motivate as well as enable users to behave secure within the organizational environment. The wave cannot be assigned to another shift of computing paradigms having taken place and the instrument of organizational culture can also be interpreted as "just" another tool of managerial practice. It is thus reasonable to subsume von Solms' second and third "Wave" under a common "Era" of distributed systems, where managerial aspects were rated significantly more important for information security.

Conclusion:
The "Third Wave" should, together with the "Second Wave", be subsumed under a *common "Era" of distributed systems*, which was shaped by decentralization of computer usage, growing user flexibility and rising user responsibility. This "Era" can still be characterized as an era of human-related and managerial approaches.

2.1.4 Spanning Discussion – from Waves to Eras

We have so far described and discussed the "Three Waves of Information Security" as introduced by von Solms in 2000. It can be argued that the model mainly has a descriptive nature. Von Solms observes different changes of dominating approaches having "happened" within the field of information security over time. These changes of approach are then condensed by time of occurrence and subsumed under the abstract term of different "Waves" having shaped the way of thinking about information security during the respective times. Even if this perspective helps a lot in structuring the different areas of information security and is thus indeed valuable, von Solms' model still reveals some incompletenesses and inconsistencies that have been identified within our discussion.

[31] The underlying distinction of ex-ante and ex-post approaches will also be discussed in section 2.2 and, more abstract, in chapter 7

[32] On the other hand, it could be argued that the instruments differ too much from each other to be treated as the same. Section 2.2 follows this distinction and gives a couple of arguments for doing so. Perhaps instructions and security culture should be assigned to different phases of the same "Wave" or "Era".

First, the "Era" of *isolated computing* is completely ignored. Information security was already important during this era and, as mentioned above (see p. 31), was mainly ensured by means of *physical protection*. Even if von Solms begins his discussion with the *mainframe-era*, which led to *technical approaches* for information security, the era of isolated computing and the corresponding approach of physical protection should be included in a well-founded model of information security based on historic succession.

Furthermore, von Solms identifies the "Second Wave" by the emergence of information security officers and organizational structures. With a purely descriptive intention, this is of course warrantable, but for identifying methodically different approaches to information security, this is hardly relevant. Organizational structures could also have been erected to manage the established approaches in a different way. Thus, the introduction of *information security guidelines* being derived from information security policies has been identified as the main factor for the identification of a new, "Second Wave". This can be carried over to a third "Era" of *distributed computing*.

And third, it can be discussed whether the "Third Wave" as proposed by von Solms should actually be treated as such. Two waves, the second and the third one, refer to managerial aspects of Information security: The "Second Wave" is mainly shaped by law-like policies and the third one by aspects of organizational culture, which are both well known as managerial tools. Unquestionably, the approaches still differ from each other and it could thus be argued that the second and the third wave should be treated as two different parts of a common "Managerial Era".

Additionally, von Solms does *not* identify, propose or even search for something like an overarching mechanism or model – transcending historic succession – that explains and reasons the unquestionably existing expansions in thinking about information security and in implementing it.

As it can be seen from the above discussion, the way of approaching information security and thus the emergence of new "Eras" was shaped by the different prevailing computing paradigms: Physical protection was the dominant approach during the "era of isolated computing", the "mainframe-era" led to the development of additional technical instruments and the "minicomputer-" or "PC-era" to a more managerial approach that considered human-related means.[33] During this ongoing change, it can be observed that the different processes of storing and modifying information more and more moved away from central instances and were increasingly placed on distributed computers locally. This process of the "decentralization of computing" and the respective widening of approaches for information security are illustrated in figure 2.2.

With this model, we have proposed a first classification system for the different existing measures based on their historical emergence. Additionally, the different types of measures can be associated to major shifts having changed the prevailing way of using computers inside organizations. What this model achieves is *explaining* observable changes having taken place *in the past*. It also helps, as well as the underlying

[33] A comparable taxonomy for distinguishing different existing approaches was also given by Bishop (2003, p. 93), distinguishing physical, technical and procedural mechanisms. Koops (1999, pp. 26 f) distinguishes *"physical, technological, and organizational measures"*.

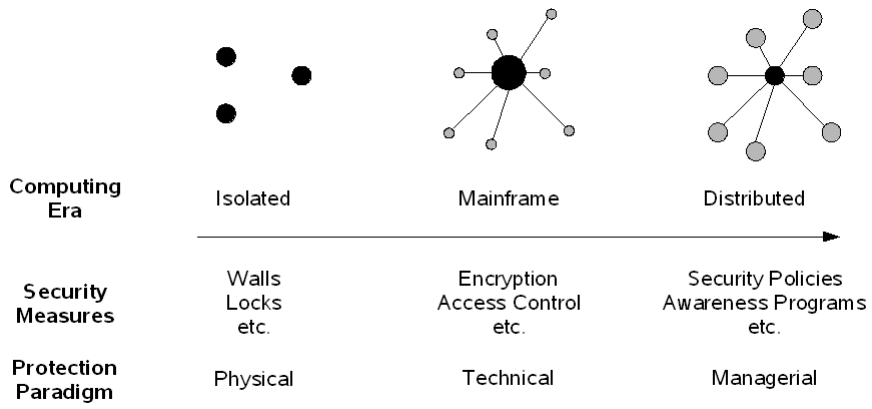


Figure 2.2: Decentralization in computer usage and change in security approaches

model by von Solms does, to structure considerations.

But the model does so far not support discussions about *further* extensions such as possible new kinds of approaches which could help dealing with information security problems that emerge currently or even in future. In short: The predictive capabilities of the developed model are poor. Additionally, historic succession is of course not the only possible dimension for structuring information security approaches. Other models based on different dimensions may provide better expandability or at least suggest a direction where to search for further improvements. In the following section, we will therefore consider an alternative model.

2.2 The Security Triangle

The previous section introduced a model for structuring the field of information security inside organizations on a historical basis. In this section, we will derive another model. This model is partially similar to the model of the "Three Waves" considered above and the derived model of the "three eras" presented in section 2.1.4. Nonetheless, the model mainly differs from the aforementioned one in the taken point of view. It classifies the different existing security measures by their character and subsumes them under three distinct "meta-measures" representing generalized types. Thus, the model takes the *nature* of the different existing instruments as encompassing dimension. These "meta-measures" can be characterized on the basis of three different properties:

- **Way of implementation:** This property is the most important one for distinguishing the separate classes. Considering the way a certain instrument is implemented allows to distinguish different types of measures and thereby to identify the different "meta-measures". If a certain measure is mainly shaped by the use of technical artifacts, this clearly distinguishes it from a measure that is primarily based on written documents. Nonetheless, this property is rather

descriptive than analytic.

- **Strictness:** This property has an analytic purpose and stands for the accurateness and the explicitness of a certain measure or "meta-measure". An instrument that clearly differentiates forbidden actions from allowed ones is thereby clearly distinguished from another instrument that is based on less well-defined rules. Additionally, the strictness of a certain measure also refers to its flexibility. An instrument that provides no possibilities for instant reaction to exceptional and unforeseen situations is within here considered more strict than a measure that intentionally allows overriding. An example from the non-IT world might illustrate this aspect: A barrier that is used to stop cars does – in contrast to red traffic lights – not provide a possibility to disregard it in case of an emergency. This barrier would therefore be identified as being more strict than the traffic lights.

- **Enforcement approach**: This property distinguishes the different measures by their mode of enforcement. Mainly, this distinction is between the ex-ante and the ex-post approach. To use the above example again, a traffic light follows the ex-post approach which is based on the menace of later punishment and thus mainly works through deterrence, while a barrier uses the ex-ante approach by making it impossible to pass at all. For the field of information security, Straub (1990) distinguished "deterrents" from "preventives", where "deterrrents" can be mapped to those means using an ex-post approach and "preventives" to those being shaped by an ex-ante approach of enforcement.

Based on these properties, three "meta-measures" currently being used for information security inside organizations can be distinguished. These are *architectural means*, *formal rules* and *informal rules* and they are characterized as follows:

- **Architectural means** include the *physical* and the *technical* means as already outlined in section 2.1.1. These instruments share some common characteristics: First, they employ *artifacts* – physical or nonphysical ones – to accomplish a certain goal. Second, they are *strict*, which means that well-defined rules are (or at least can be) expressed through them and that they do not arrange for reacting to exceptional situations that have not been foreseen during their setup. And third, architectural means follow an *ex-ante approach*, which means that unwanted activities are tried to be prevented *before* taking place at all.

- **Formal Rules** refer to means that are not focused on technology but rather accurately define which activities are obligatory, which ones are allowed and which ones are forbidden. An example for such formal rules are the guidelines mentioned in chapter 2.1.2.2 which will be examined in more detail below. Formal rules can also be identified by three significant characteristics: They are – different from architectural means – realized through *written words* instead of technical artifacts. They are *relatively strict* because they are comparably well-defined and because the ability of taking exceptional situations into account is

existing but certainly low. And finally, formal rules follow an approach of *ex-post enforcement*, which means that forbidden activities are not made impossible but rather punished after having happened.[34]

- **Informal Rules** refer to those rather vague factors influencing individual behavior that are usually associated with a *"security culture"* and that do, to a certain extent, specify what constitutes "good behavior" in a rather social manner. Some of these rules can be written down – in a code of conduct, for example – but the comprehensive body of informal rules always includes more than written statements. Informal rules are *less strict* than the aforementioned formal ones, because they do not exactly define what behavior is acceptable and what not. In return, informal rules make it much easier to respond to unanticipated events. And finally, enforcement of informal rules is more complex than it is for architectural means and for formal rules. Basically, however, enforcement is realized through social (instead of hierarchical) control and *ex-post enforcement*.

Besides these "meta-measures", there also exists another class of means that will not be considered for the development of the alternative model for several reasons, even if it gets much attention in the literature. Nonetheless, it should be mentioned and illustrated in short to explain its non-consideration:

- **Organizational support** refers to instruments that affect information security not directly but rather indirectly. A good example for such means of organizational support is the establishment of specific management structures for information security like chief information security officers (CISO's). The importance of such dedicated management structures being erected is undoubted in the literature as well as in practitioner's world. But management structures are no security means by themselves. Rather, they support the planning and the implementation of security-related means inside the organization and thus care for a professional and structured way of dealing with information security in general. Means of organizational support are for this reason not considered as instruments *directly* affecting information security and are thus not included within our model of the "meta-measures". Their importance, however, shall in no way be called into question.

The above-mentioned three "meta-measures" form the general set of possibilities present to an organization for approaching information security. Let us therefore consider these "meta-measures" and their distinct characteristics in some more detail to support the development of our positive model of information security inside organizations.

[34] This mechanism allows for a certain possibility of reacting to exceptional situations by abandoning punishment or by varying the degree of penalty. Nonetheless, the underlying model is still a rather strict one in the aforementioned sense.

2.2.1 Architectural Means

The "meta-measure" of architectural means refers to physical as well as technical artifacts being used to enhance information security. These "artifacts" include means like those mentioned in chapter 2.1.1 (Thick walls or sophisticated locks as means of physical protection and encryption or access restrictions as technical means) but are not limited to them. Over time, technical means have been developed and enhanced continuously to meet new needs and challenges. For example, instruments like firewalls or DRM-systems also belong to the category of architectural means.

As used herein, the term of architectural means includes all human-made objects directly being used for the purpose of information security. These objects can be physical – like constructional means – as well as nonphysical – like firewalls being realized as a software program. Architectural means thus represent the traditional, technology-focused view on information security. Even if it is nowadays widely accepted as being out-dated to concentrate on this type of measures exclusively, architectural means still play an important role as an integral part of what is usually called a "holistic approach" to information security.

2.2.1.1 Strictness

Architectural means feature, on a generalized basis, the highest degree of strictness from the three mentioned "meta-measures". In contrast to all other approaches, the underlying model is basically a binary one. Take, for example, doors, access control mechanisms or firewalls: Admittance to a data center either is allowed through key-ownership or it is not. Access to a certain file either is allowed by access rights or it is not. And an outbound network connection on a certain port either is allowed or it is blocked. In all these cases, the rules are defined explicitly and accurately and there exists no space for interpretation.

The strictness of architectural means primarily originates from their nature of being based on artifacts. Artifacts are only able to exactly put those rules into action that were implemented into them before and the rules that should be enforced by the artifact have to be formulated explicitly and in a discrete manner as the artifacts only operate on a basis of "yes" and "no", without any possibility for trade-off. In contrast to other measures, physical and technical artifacts do not provide any mechanisms for "reasonably" breaking the well-defined rules and are thus *highly strict*.

2.2.1.2 Ex-ante Enforcement

In most cases, architectural means are used to heighten information security through restrictions that limit the users' possibilities to a well-defined set of opportunities and thereby prevent unwanted actions *before* taking place.[35] The above-mentioned fact of architectural means being the most "strict" of the three meta-measures makes this possible: As regulations regarding a certain action are clear, unambiguous and

[35] See, for example, Reidenberg (1998, p. 568, note 106): *"Technology may, however, prevent an action that violates the rule from occurring at all."*

without space for interpretation, decisions on the admissibility of a certain action are considered deterministic and can therefore easily be "delegated" to artifacts like software programs or even buildings.

This delegation is usually realized by *implementing* the regulations *into* the artifacts – the ways for doing so are manifold and include software configuration as well as architectural choices – and by "equipping" the artifacts with an enforcement mechanism. Even if this enforcement mechanisms could also be realized on an ex-post basis, for example by reporting identified rule violations to supervisors, ex-ante enforcement is far more common. And due to the unambiguous nature of the implemented rules, this practice is reasonable: If rules are well-defined and clear, if there is no doubt that a certain user must not have access to a certain file, then it seems to be much wiser to prevent access *before* it happens instead of punishing the user for unauthorized access later. Of course, this puts strong importance on the rules being *really well-defined and clear*, but this does not affect the underlying nature of the "meta-measure" of architectural means. This is, in any case, significantly shaped by the model of *ex-ante enforcement*.

2.2.1.3 Additional Remarks

Besides those approaches to information security that are aimed at restricting the users' choices, there also exist technologies that do not restrict anyone but rather provide new opportunities of doing the same things in a more secure manner. Redundant hardware falls into this category: It provides higher security – as it heightens availability – without entailing any constraints. Such means could be called *"non-restricting architectural means"*. Even if such a model of "invisible" and "interference free" security is usually one of the primary goals of security efforts, it can only be achieved in very few cases. But if such an instrument can be identified, there would be no reason – possibly besides its cost – not to make use of it. Thus, "non-restricting architectural means" will presumably not be the subject of any considerations besides monetary ones. Therefrom, they are effectively irrelevant for further considerations.[36]

Conclusion:
The "meta-measure" of *architectural means* for information security can be characterized by the use of *physical or technical artifacts*. Architectural means regulate on an *ex-ante* basis, are *well-defined and clear* and can be considered *highly strict*.

[36] The reader shall be aware that really interference-free and non-restricting means are absolutely rare. For example, smartcard-based login procedures could be seen as an interference-free enhancement of security. But if the use of smartcards is obligatory, it is impossible to logon without the smartcard in case of an emergency. There are many other possible examples for mechanisms being non-restrictive only at first sight.

2.2.2 Formal Rules

Formal rules were described above as written words accurately defining what activities are obligatory, which ones are allowed and which ones are forbidden. The best example for formal rules regarding information security are written information security policies. Wiant (2005, p. 453) distinguishes *"explicit policies"* from *"implicit"* ones, where *"explicit policies provide an official focal point for the entire organization"* and *"will generally be codified in a written form"*. Implicit policies, in turn, are classified by Wiant as *"less formal than explicit ones"*. Due to this less formal character, implicit policies will be imputed to the "informal rules" mentioned below.[37]

Additionally, Karyda et al. (2005) distinguish the term *"security policies"* and the term *"security guidelines"* with a guideline being part of or derived from a more comprehensive policy, which also includes technical measures and other instruments. Thus, to use clear and unambiguous terms, the "meta-measure" of formal rules only refers to the explicit part of what Karyda et al. identify as "security guidelines".[38]

Examples for such explicit security guidelines can be found in a multitude of sources. We will here consider one of the policy templates provided by the SANS Security Policy Project[39] for demonstrative purposes: The "Email Use Policy" (SANS 2006). Even if this policy does not directly refer to information security aspects, it was chosen because it illustrates the different aspects of formal rules considered below very well.

2.2.2.1 Strictness

As mentioned above, the "strictness" of a certain measure refers to its accurateness and explicitness and a strict measure represents well-defined rules regarding allowed and forbidden activities. Security guidelines, which are considered here, are, according to Karyda et al. (2005, p. 248), *"of prescriptive nature, constituting an imperative for [information system] users."* This prescriptive nature is a sign of strictness and it can also be found in the aforementioned "Email Use Policy":

> *"3.1 Prohibited Use.*
> *The ⟨COMPANY NAME⟩ email system shall not to be used for the creation or distribution of any disruptive or offensive messages, including offensive comments about race, gender, hair color, disabilities, age, sexual orientation, pornography, religious beliefs and practice, political beliefs, or national origin. Employees who receive any emails with this content from any ⟨COMPANY NAME⟩ employee should report the matter to their supervisor immediately." (SANS 2006)*

This short example, which can be considered representative for the "meta-measure" of formal rules, illustrates their strict character quite well. It is relatively accurate, as it explicitly lists a couple of definitely unacceptable kinds of messages being sent over the company's email system. Nonetheless, it is not as entirely well-defined as it is the

[37] See section 2.2.3
[38] For further comments on the distinction of policies and guidelines, see also section 2.1.2.2.
[39] The project's website can be found at `http://www.sans.org/resources/policies/` [05.10.2006].

case for architectural means, because it also prohibits other *"disruptive or offensive messages"* not being explicitly mentioned.[40]

As this exemplifies, the fact of written language being used as medium also allows for formal rules that are *intentionally* non-strict to a certain extent. Take, for example, another point from the mentioned "Email Use Policy":

> *"3.2 Personal Use*
> *Using a reasonable amount of ⟨COMPANY NAME⟩ resources for personal emails is acceptable, but nonwork related email shall be saved in a separate folder from work related email. [...]" (SANS 2006)*

The conscious use of the interpretable phrase *"reasonable"* is a possibility only natural language offers. By using such a phrase, the strictness of formal rules can be lessened intentionally as different employees will understand different amounts of usage being *"reasonable"*.

As it becomes clear, formal rules do not have one defined level of strictness. On the one hand, they can be used to explicitly and accurately specify which actions are obligatory, which ones are allowed and which ones are forbidden. Such formal rules can be called highly strict.[41] But on the other hand, formal rules can – through the conscious use of written language – also be enriched with a well-dosed degree of non-strictness. At the same time, highly explicit and accurate written rules lower the possibilities for reaction to unforeseen situations. Even if formal rules do – in contrast to architectural means – generally permit their violation, such cases of disregard can be reduced by the deterrent nature of explicit statements. Non-strict statements, in turn, give people the power to break general rules if this appears to be necessary for responding to unanticipated events or situations.

Usually, organizations use both mechanisms to a certain extend. Whether this "mixture" then tends to be strict or non-strict depends, of course, on the respective organization and its particular needs. For our purpose, we will therefore consider formal rules as having a *medium to high degree of strictness*.

2.2.2.2 Ex-post Enforcement

The way of enforcement is another characteristic for distinguishing different "meta-measures". Formal rules were stated above to get enforced on an ex-post basis through punishment in case of noncompliance. A further example, again taken from the "Email Use Policy" mentioned above, illustrates the fact of ex-post enforcement being used for formal rules:

> *"4.0 Enforcement*
> *Any employee found to have violated this policy may be subject to disciplinary action, up to and including termination of employment." (SANS 2006)*

[40] The formulation in the example would be more comprehensive if it was "... including *(but not limited to)*..."

[41] For the strictness of this kind of rules, see also Dhillon and Backhouse (2000, p. 128): *"Rules apply in foreseen and predictable circumstances and cannot be invoked in new and dynamic situations."*

As it becomes clear from this example, the method of ex-post enforcement does not provide absolute protection from unacceptable messages being sent over the company's system. Different from the ex-ante enforcement of architectural means, it does not make unwanted activities impossible in general. Instead, breaking the rule is still possible, even if being associated with a certain probability of detection[42] and, in case of detection, with a fine or another punishment. Formal rules thus have to be seen as *"deterrents"* which are enforced by means of *"disincentives"* (Straub 1990).

As stated above, the intentional use of interpretable formulations can lead to a less strict character of the respective rules. This does, however, lead to implications for the enforcement process: If an employee sends an email over the company's network that includes, for instance, offensive comments about sexual orientation, this would definitely be a breach of an explicitly defined rule and thus had to be punished. But if an employee sends and receives, say, one hundred private emails a day, it is *not definitely* clear whether this is still a *"reasonable amount"* or not. Consequentially, it is also unclear whether the respective employee has to be punished or not. The meaning of the word *"reasonable"* has to be interpreted from case to case and under consideration of additional circumstances. Informal rules, as described in section 2.2.3, could also play an important role for this evaluation.

All this, however, does not alter the general enforcement approach of formal rules, which still follows the *ex-post* model.

2.2.2.3 Additional Remarks

The establishment of an information security policy including clear instructions for the users and other affected persons – as the herein considered explicit guidelines – is usually deemed as positively affecting information security in general. This was put into question in a study conducted by Wiant (2005), which suggests that even while the existence of an information security policy leads to higher reporting rates of incidents, it does *not* lead to fewer or less serious incidents in general.[43]

Another study, conducted by Straub (1990), suggests the opposite: *"[Information systems] security deterrents result in reduced incidence of computer abuse."* Unfortunately, there are no similar studies publicly available and thus, the strength of formal rules and their real impact on information security is still unclear. But in any case, they represent an existing and well-established approach and will therefore be considered within our model of the three "meta-measures".

Conclusion:
The "meta-measure" of *formal rules* represents explicitly stated rules governing security-related behavior of an organization's members. These rules are enforced by *ex-post* sanctions and have a *medium to high level of strictness*.

[42] Interestingly, this probability is intentionally increased by the obligation to report the receipt of unacceptable messages to the supervisor mentioned in the first excerpt in section 2.2.2.1.

[43] See Wiant (2005, p. 456): *"[A]fter all the comments in the literature plus federal and state regulation that dictates policy statements be created to safeguard medical records, [...] from this study these actions do not seem to be warranted."*

2.2.3 Informal Rules

Besides the relatively clear and observable "meta-measures" of architectural means and formal rules, there also exists a set of more "informal rules" constituting what several authors have called "security awareness"[44] or a "security culture".[45] The nature of such "informal rules" regarding information security is – even after extensive literature review – fairly vague. Perhaps the best description for the aim of "informal rules" regarding information security is that they are expected to make secure behavior become *"part of habit"* (Chia et al. 2002) for the members of an organization. Different from architectural means or formal rules, such "informal rules" are neither materialized through artifacts, nor can they be written down completely or entirely represented in any other investigatable form. They are thus harder to observe and – consequently – harder to analyze. Nonetheless, they unquestionably exist and affect individual behavior in nearly every organization.

The principles of tacit consent and common behavior play a crucial role for the "meta-measure" of "informal rules". Consider, for example, an organization that allows private Internet usage but that neither uses formal rules nor architectural means to regulate which sites are acceptable to access from the organization-internal network and which ones are not. Even in this case, there would in all likelihood be a tacit consent among the organization's members that keeps them from accessing certain, say, dubious web-sites.

Another example is an organization where it is (by means of formal rules) prohibited to use other USB-devices than keyboards and mice on a workplace-PC. Even under such circumstances, it could be consent among the members that the use of external storage media is acceptable within certain situations.[46] As this example shows, there could also be a tacit consent about acceptable behaviour which is shaped by a certain set of informal rules that *contradict* the given set of formal rules.

Many further examples could be given for "informal rules" that affect human behavior with regard to information security but in any case, such "informal rules" require a certain degree of consent within the organization to be effective and to define a body of common behavior. We will discuss this unquestionably important aspect in later chapters and shall, for the moment, proceed with the characterization of "informal" rules on the basis of the same properties that were already used for the previous "meta-measures".

2.2.3.1 Non-Strictness

Regarding the dimension of strictness, informal rules are neither accurate nor are they explicit. They do not precisely define whether a certain action is allowed or not,

[44] See, for example, Siponen (2000)

[45] See, for example, Chia, Maynard, and Ruighaver (2002) or OECD (2002).

[46] Consider, for example, a representative of the organization having to take with him the latest version of a presentation when leaving for a congress.

but rather allow to make decisions in consideration of the particular context. Even if it is tacit consent within an organization that using external storage media on a workplace-PC could be deemed acceptable in case of urgency, this does not need to be true for the case of having enough time for following a more formal procedure to transfer internal data onto external media.

Thus, informal rules do not clearly differentiate between forbidden and allowed actions and regulate in a more analogous manner instead of a binary one. The rules are less well-defined than it is the case with architectural means or with formal rules and thus offer wide space for individual and situational trade-offs. Of course, this can lead to uncertainty about the acceptability of a certain action, but at the same time, it allows for reaction to exceptional situations or unanticipated events.

On the other hand, informal rules can also lead to a certain degree of strictness. For instance, it was stated above that *formal* rules can be formulated in a way that leads to decisions being dependent on the respective set of established *informal* rules.[47] If it is stated in a formal rule that a "reasonable amount" of private emails is acceptable and if the interpretation of "reasonable" has to be made on the basis of a set of established informal rules, these informal rules could – even if not explicitly – very well indicate that hundreds of private emails per day over a period of several months are unquestionably more than "reasonable". As such constellations can bring a certain amount of strictness into informal rules, they will herein be considered as being shaped by *low to medium strictness*.

2.2.3.2 Ex-post Enforcement

Informal rules do not avoid unaccepted behavior ex-ante and thus generally allow infringements. But at the same time, there also are no formally defined procedures for sanctioning unaccepted behavior, which could be argued against the identification of an ex-post enforcement model for the "meta-measure" of informal rules.

In fact, however, the enforcement model of informal rules differs from the model of the aforementioned "meta-measures" in a pivotal aspect. For architectural means as well as for formal rules, enforcement is asserted by some dedicated singular instance that is assumed to hierarchically stand above the objects of sanction. Architectural means are implemented, for example, by facility management and IT departments on behalf of and with backing from the organization's executives whilst formal rules are enforced by the respective supervisors or through some other kind of intra-organizational disciplinary authority. Thus, enforcement is in both cases realized on the basis of a hierarchical model where some kind of authority has power over those who are ultimately affected by the respective instrument.

This is different for informal rules. It was mentioned above that informal rules are based on a body of tacit consent among the members of an organization and thereby form some kind of common behavior. The principle of decentralization plays an important role for such constellations. Tacit consent cannot be "ordered hierarchically"

[47] See section 2.2.2.1

but rather has to evolve and establish within the respective group of people.[48] The often stressed term of a "security culture" also emphasizes this view.

Enforcement of such informal rules works similarly. The idea behind this mode of influencing individual behavior is based on the principle of *"informal control"* (Ellickson 1991, p. 131). In this model, every member of the organization supervises – to a certain extent – the behavior of other members and objurgates them in case of an infringement of an established informal rule. Other sanctions beyond objurgation are also possible. For example, a member could reject working together with another member because of knowing him not to have followed well-established informal rules *in the past*.

Altogether, enforcement of informal rules has two main characteristics: First, informal rules are enforced on an *ex-post* basis. Unaccepted or rule-breaking behavior is not prevented from happening at all but rather sanctioned after having taken place. And second, enforcement of informal rules is realized in an *informal, decentralized manner* instead of the strongly hierarchical way of sanction which is employed for architectural means and formal rules.

Conclusion:
Informal rules form the third "meta-measure" affecting information security inside organizations. They evolve from a *tacit consent and common behavior* among the members of an organization. Informal rules regulate through *ex-post* mechanisms and are of *low to medium strictness*.

2.2.4 The Regulatory Framework as "Field of the Game"

Of course, the identified meta-measures do not exist without a higher-level context which has to be considered during the analysis of the current status quo of information security, too. For realizing information security, no organization has completely free choice of means. Instead, every organization faces a couple of external regulatory constraints defining possibilities, restrictions and obligations regarding information security and the way it is addressed.

Take, for instance, the enforcement of formal rules mentioned above (section 2.2.2): Even if it would seem reasonable for an organization to enforce such rules by threat of draconian punishments, this would in most cases presumably be prohibited by diverse legal restrictions.[49] The same is true for some kinds of monitoring employees compliance with formal rules: Legislation influences an organization's choice of means here, too.

[48] Nonetheless, the evolvement of informal rules can very well be *influenced* through hierarchical means. For example, members of hierarchical bodies have the authority to brief their subordinates or to appeal to them in different ways. In this way, tacit consent can at least be biased through hierarchical structures. And in fact, this is the way that is often proposed for consciously forming a "security culture" inside organizations from the executive level downwards. We will discuss this aspect in more detail in section 7.2.3.

[49] In fact, even the punishment being mentioned above (*"up to and including termination of employment"*) would be considered disproportional for most breaches under certain legal regimes, especially in Europe.

On the other hand, the "regulatory framework" not only circumvents the use of measures that would otherwise possibly be considered expedient by the organization. Some laws also force organizations – as soon as they are affected by the respective law – to use mechanisms that would not be put in place from the organization's interest alone. Privacy laws are a good example: The Health Insurance Portability and Accountability Act of 1996 (HIPAA 1996) as well as the European Privacy Directive (European Union 2002) and its national implementations prescribe certain sets of measures that have to be implemented by the affected organizations even if these organizations have no own interest in doing so.

The Sarbanes-Oxley Act of 2002 (SOX 2002) also got much attention from information security professionals over the past few years.[50] The act itself does not include tangible prescriptions on how information security has to be implemented within the affected organizations but rather requires the management to establish and maintain *"an adequate internal control structure and procedures for financial reporting."* (SOX 2002, section 404) and to *"certify financial statements and the existence and effective operation of disclosure controls and procedures"* (Damianides 2005, p. 78). These obligations alone would not restrict an organizations choice of means too strongly and thereby leave it to the organization how to fulfill the requirements.

Nonetheless, it is usual practice to prove compliance to SOX by verifying compliance to different information security standards or frameworks.[51] For SOX, the "Public Company Accounting Oversight Board" explicitly mentions COSO (1994) as being a *"suitable framework"* for verification of compliance[52], but it is also possible to use other "suitable" frameworks. Pinder (2006, p. 34), for example, mentions CobiT (ITGI 2007) as an accepted "suitable" framework, too. Even if the Sarbanes-Oxley-Act itself does not prescribe in detail how information security has to be managed inside an organization affected by the act, the respective organizations are nonetheless – through the explicit mention of existing security frameworks as being "suitable" – indirectly forced by the act to comply to one of these frameworks. The Sarbanes-Oxley-Act thus also limits an affected organization's options for choosing an own way for realizing

[50] The Act was established after several financial scandals in the US and its main purpose lies in the financial area rather than in the field of information technology. Nonetheless, it has strong impact on the management of information security inside organizations. Sections 302 and 404 of the act are most frequently mentioned in conjunction with information security whilst Kaarst-Brown and Kelly (2005) also mention section 409 as important with regard to information security. In any case, *"many of the other sections also have implications for the IT function"* (Kaarst-Brown and Kelly 2005, p. 1). The complete identification of SOX-sections relevant to information security is not necessary here, as we only consider SOX in order to highlight the *existence* of an external "regulatory framework" that has to be respected by organizations.

[51] This is also the case for other legislative rules. For example, regarding the HIPAA mentioned above, the SANS Security Policy Project mentions *"18 information security standards in three areas that must be met to ensure compliance with the HIPAA Security Rule"*. See http://www.sans.org/resources/policies/ [16.02.2009]

[52] See PCAOB (2004, p. 8): *"COSO's publication (also referred to simply as COSO) provides a suitable framework for purposes of management's assessment."* See also PCAOB (2007, p. A1-6, note 7): *"SEC rules require management to base its evaluation [...] on a suitable, recognized control framework [...]. For example, the report of the Committee of Sponsoring Organizations of the Treadway Commission (known as the COSO report) provides such a framework [...]."*

information security.[53]

Additionally, the example of SOX shows that established standards like COSO, CobiT or ISO / IEC 27001 (ISO / IEC 2005b) have – besides legal regulations – also to be considered as part of the "regulatory framework". The wide range of existing standards shall not be discussed in detail here.[54] For our purposes, it is enough to be aware of their existence and of the fact that compliance to them does not always take place voluntarily because of internal considerations – for example to enhance the quality of the internal incident management – but rather can be obligatory or quasi-obligatory for organizations, too.

For the cases mentioned above, compliance to standards gets obligatory for those organizations being affected by a certain law. The quasi-obligatory case, in turn, occurs when there is no legal obligation for compliance but when compliance is necessary for business reasons. Contractors can, for example, require compliance to a certain standard for starting business connections or for the permission to participate in an invitation to tender, like it is often the case for contracting with authorities. Pinder (2006, p. 32) also mentions "SOX by proxy" which exactly represents this situation (even with an extra loop through a legal directive): Organizations being obligated to comply to a law and thereby to a standard can also require *"its suppliers to demonstrate a similar degree of internal control"* (id.) and thereby make standard-compliance quasi-obligatory.[55]

Altogether, the "regulatory framework" – as understood herein – consists of all external constraints limiting an organization's freedom of choice for the realization of information security. As it was shown, it contains at least (but is not limited to) *laws* and *established standards*. Both of these externally influence the way information security is implemented inside organizations. They limit organizations' choice of means by forcing them to use certain instruments they would not implement otherwise and by restricting the utilization of means that would seem reasonable in certain situations.

The "regulatory framework" thus clearly constrains an organization's possibilities for addressing information security. Every strategy, every approach and every activity has to respect these boundaries. The "regulatory framework" therefore defines what can

[53] The author is aware of the vital discussion regarding the actual effect of the Sarbanes-Oxley Act for information security and even for the original aim of fraud-prevention. See especially Ghose and Rajan (2006), suggesting that small firms are handicapped in comparison to larger ones due to the one-solution-fits-all approach of SOX and similar regulations. See also Anderson and Moore (2007, p. 15), concluding that *"mandatory investment in security compliance can create unintended consequences from distorting security markets to reducing competition."* All this notwithstanding, the act currently has to be taken as fact and thereby forms a part of what is herein called the "regulatory framework".

[54] A German publication by the national Association for Information Technology, Telecommunications and New Media and the national norming institute, for example, mentions 24 relevant standards for managing and evaluating information security alone. Notabene, this does not include specific standards on physical protection, cryptography etc. See BITKOM and DIN (2006)

[55] As Cavusoglu et al. (2005) point out, this effect can not only occur with explicit *"coercive"* pressure from other parties (other organizations, the state, etc.) but on the pure basis of *"mimetic"* and *"normative pressures"*, too. Their concept of legitimacy-based reasons for organizations to behave in a specific manner with regard to information security can perfectly be attributed to the concept of the "regulatory framework" as used herein. See also Shostack and Stewart (2008, p. 114).

be called "the field for playing the game of information security" inside organizations.

Conclusion:
Besides the three meta-measures, there exists a *regulatory framework* – consisting at least of laws and established standards – that limits an organization's choice of means. This regulatory framework sets certain boundaries for developing an organization-internal strategy for realizing information security.

2.2.5 The Security Triangle - Synopsis

We have so far discussed three currently established meta-measures for realizing information security inside organizations: architectural means, formal rules and informal rules.

Architectural means are realized by the utilization of artifacts – buildings as well as IT systems – and regulate security-related behavior of an organization's members on an *ex-ante* basis. They are *highly strict*, which means that they are well-defined and explicit and do not give users the possibility to disregard them even in exceptional situations.

Formal rules stand for explicit guidelines being obligatory for all members of an organization and are usually part of a more comprehensive security policy. Such guidelines affect the members' behavior on an *ex-post* basis which means that breaching the rule is generally possible but can be sanctioned after the fact. Additionally, formal rules feature a *medium to high strictness*, because some formal rules are clear and well defined and offer no possibility for interpretation or trade-off while other formal rules are consciously vague and cannot be judged without interpretation and consideration of the respective context.

And finally, informal rules regulate security-related behavior, too. They also act by *ex-post* enforcement as unacceptable activities are not prevented but rather sanctioned after having happened. Different from the other meta-measures, sanctions for informal rules are not realized through centralized authorities from some kind of hierarchical body but rather in a more decentralized process of informal control. Informal rules have – especially in comparison to the other meta-measures – a *low to medium level of strictness*. They cannot be articulated or represented entirely and are thus neither well-defined nor are they explicit. Consequently, interpretation and consideration of context play a crucial role for the meta-measure of informal rules.

Table 2.1 summarizes the main properties of the three identified meta-measures. These meta-measures have so far been considered independently from each other, but there unquestionably are strong interdependencies between them. Each of the meta-measures influences information security directly as well as it influences the other meta-measures, too.

Take, for example, architectural means. If an organization introduces a technical system that prevents any outbound network connection aside from from a set of well-defined exceptions, it would be obsolete to have formal rules explicitly specifying which applications using external connections are allowed and which ones are not. Similarly,

Table 2.1: Three meta-measures for information security – Main properties

Meta-Measure	Enforcement Approach	Strictness
Architectural Means	Ex-ante	High
Formal Rules	Ex-post	Medium – High
Informal Rules	Ex-post	Low – Medium

a technical system that is introduced to prevent using different external devices than keyboards and mice could result in strong obstructions for daily practice and thereby lead to the emergence of a shared consent amongst the members (an informal rule) that circumventing the system actively and consciously is acceptable under certain circumstances.

The same is true for the other meta-measures. A formal rule explicitly prohibiting to send emails including offensive comments about national origin could – e. g. through sensitization – give rise to a shared consent that comments about *regional* origin are also unacceptable. An established informal rule regarding the acceptability of websites being accessed from the internal network could make technical systems for white- or blacklisting websites unnecessary, etc. Interdependencies between meta-measures are rather the normal case than the exception.

Additionally, section 2.2.4 identified a higher-level regulatory framework influencing the way an organization can make use of the three different meta-measures. This regulatory framework includes at least legal regulations and established standards limiting an organization's freedom of choice for implementing information security. Certain practices can be prescribed by laws, others can be prohibited even if seeming reasonable and compliance to standards can be obligatory or quasi-obligatory even if the standard does not meet the specific needs of an organization. The regulatory framework thus forms the "field of the game" for making use of the different meta-measures inside organizations.

Altogether, we have so far outlined a model consisting of three meta-measures for realizing information security inside organizations which can only be used within certain boundaries given by a regulatory framework. The meta-measures have been examined separately to identify some distinguishing main characteristics, but in real contexts, there will always be interdependencies between them which have to be considered, too. This leads us to the graphical representation of the model shown in figure 2.3. This model will from here on be called the "security triangle".

This model represents an abstract structure of the different means that are currently used within the field of organizational information security. It can therefore be used to heighten understanding of the different approaches currently being discussed as well as for explaining the interdependencies between them. Additionally, the model offers possibilities for identifying potential shortcomings of the currently established set of approaches. Besides the model of the different "eras" from section 2.1, we will therefore use the "security triangle" as further foundation for developing our positive model of

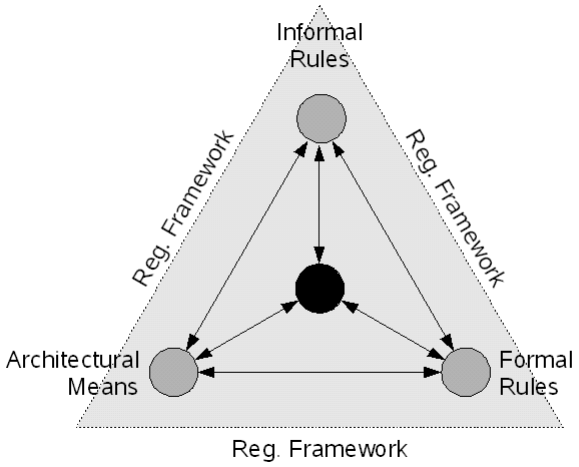


Figure 2.3: The "Security Triangle"

organization-internal information security.

Conclusion:
Information security inside organizations is currently realized on the basis of three meta-measures: *architectural means, formal rules and informal rules.* These meta-measures are not isolated but rather influence each other. An organization's possibilities of making use of these meta-measures are restricted by a superior *regulatory framework* including laws and established standards. Altogether, the meta-measures and the regulatory framework form the *security triangle.*

2.3 Conclusion

In this chapter, we gave an overview of the management of information security inside organizations. We introduced two different models that structure the field from two different points of view. Section 2.1 uses the historical dimension and recapitulates the model of the "Three Waves of Information Security" as introduced by von Solms (2000). Based on a critical discussion of this model, we developed the concept of "three eras of information security" which puts major changes of prevailing information security approaches in conjunction with major changes of the predominating computing paradigm being used in organizations (section 2.1.4).

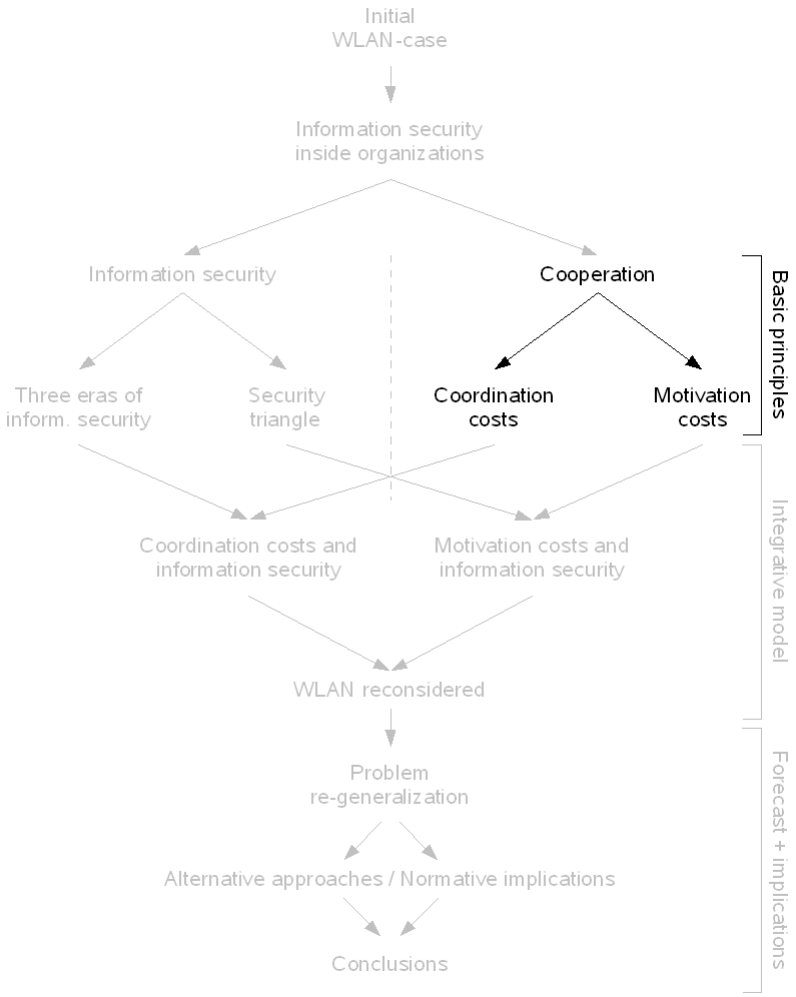
In section 2.2, we developed the model of a "security triangle" that distinguishes three "meta-measures" based on different characteristics of existing and established means of information security. Specific properties of the meta-measures were identified and the possibility of interrelations between them was noted. Additionally, we discussed the role of a superior "regulatory framework" that restricts an organization's

choice of means for realizing information security.

So far, we thus laid the foundations for our positive model from the perspective of *information security* and thereby made the first steps toward an abstract understanding of information security inside organizations. However, this is by far not sufficient for developing our aspired model. In fact, we also have to consider some aspects from economic theory in order to establish a viable, theory-founded model of information security *inside organizations*. The following chapter will therefore address these aspects in brief to allow for well-founded subsequent considerations.

Chapter 3

Some General Aspects of Organizations



Chapter 3

Some General Aspects of Organizations

> *Not the most correct, but the least complicated theories find practical application.*
>
> – Konrad Zuse

To allow for a well-founded view onto information security and its role *inside organizations*, concentrating on the "information security" aspect alone would be short-sighted. Instead, some consolidated knowledge about the nature of organizations and about principles and mechanisms forming the basis for their existence is also indispensable for substantiated discussions on information security inside organizations. This chapter thus presents – mainly from the perspective of new institutional economics – the most important fundamentals of organizational theory. It is expressly *not* the aim of this chapter to provide a detailed and all-encompassing treatise on the economic theory of organizations. Rather, only the most important principles will be examined in brief to serve later considerations.[1]

Organizations are herein, according to North (1990, p. 5), understood as *"groups of individuals bound by some common purpose to achieve objectives."*[2] This definition does not only include economic bodies like firms or trade unions but also political, social and educational bodies (id.). The emergence and existence of such organizations can be ascribed to two fundamental economic principles.[3]

The first of these principles has been described by Adam Smith (1776): Labor division and specialization allow for a more productive use of people's working power. With two people specializing on certain parts of the production of a specific good,

[1] Those readers being interested in more details of the new institutional economics of organizations may want to refer to further literature by Williamson (1985), Milgrom and Roberts (1992), Bowles (2004), Furubotn and Richter (2005), Ménard and Shirley (2005) or Picot, Dietl, and Franck (2005).

[2] Furubotn and Richter (2005, p. 296) give a comparable definition: *"[O]rganizations are generally understood as socially structured groups of individuals who seek to achieve common goals."*

[3] Other viewpoints like the interpretation of organizations as "social and cultural systems" (Scott 2001, p. xx) are not further considered herein. Nonetheless, these viewpoints shall not be refused in general. In fact, they might be of high interest for future work in the area of organizational information security. See, for example, Björck (2004). For the application of established cultural models to the field of organizational information security, see especially Glaser (2009). But for the scope of this work, organizations are primarily understood in an economic fashion.

they can together produce more units of this good than they would have been able to with each one producing the whole good on his own.[4] However, to be able to profit from this increased productivity, people (being specialized in a certain part of the production) necessarily have to *cooperate* with others during the production of the good.

Any mechanism for realizing this cooperation has, according to Milgrom and Roberts (1992, pp. 25 ff), to solve two main tasks: coordination and motivation. Of these two, the task of coordination refers to the problem of identifying which activities should be conducted by which players and how other resources should be allocated to achieve an efficient outcome. The second task of motivation refers to the problem of making people actually engage in those activities that were "assigned" to them during the previous step of coordination. According to the dichotomy of tasks having to be solved, a corresponding dichotomy of costs arising in any mechanism for realizing cooperation can be identified: coordination costs and motivation costs (see figure 3.1).[5] Furthermore, two idealized structural models exist for addressing the two tasks of coordination and motivation: markets and hierarchies. As we will see in the following, these idealized models differ significantly in their structures of coordination and motivation costs.

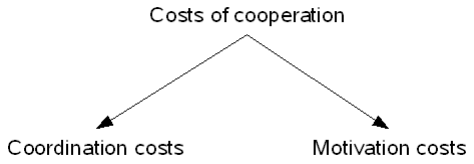


Figure 3.1: Costs of cooperation – general model

The question whether cooperation among individuals should be realized via market mechanisms – through selling half-finished products, for example – or within organizations of people being bound to each other was addressed by Coase (1937). He ascribes the emergence and existence of firms to the presence of "transaction costs"[6] and thereby introduces the second fundamental principle for the existence of firms.

Of course, this mechanism can also be applied to other kinds of organizations than firms and it will thus be addressed in more detail in section 3.1. On the other hand, any organization also faces a certain amount of costs that arise from the organizational structures themselves and from the respective constellations among different members of the organization. These costs can be summarized as the "costs of organizedness" and will be discussed in section 3.2. Together, these different kinds of costs strongly influence the structure of organizations and the way decisions are made therein. This interrelation is considered in more detail in section 3.3. And finally, ongoing technological change also changes the respective cost structures and can thus

[4] See Smith (1776, p. 110): *"The division of labour [...] occasions, in every art, a proportionable increase of the productive powers of labour."*

[5] See Milgrom and Roberts (1992, pp. 29 f).

[6] Coase himself used the term of "marketing costs" in 1937. Nonetheless, he adopted the term of "transaction costs" later (Coase 1960, see).

impact organizational structures significantly. This aspect will be addressed in section 3.4.

3.1 Transaction Costs and the Nature of Organizations

As mentioned above, the principles of labor division and specialization are fundamental for efficient value generation among individual actors. Through diversification of activities it is possible for the entirety of individuals to generate higher values than it would be possible with each individual acting independently and performing all activities on its own (Smith 1776). The *"invisible hand"*-metaphor introduced by Smith argues that through individual self-interest any member of the entirety does – under conditions of perfect market exchange – not only serve his own interests but also generates value for the entirety, thereby serving collective interests, too. Ricardo (1817) examined this principle in more detail and introduced the model of *comparative advantage* to explain why it is economically advantageous for any country to engage in the production of the good it can produce best and then exchange this good against others (produced by other countries) through market mechanisms. Obviously, this supports Smith's argumentation for labor division and specialization.

Ricardo's argument can, however, not only be applied to countries but also to individuals, like it is done in many economic textbooks.[7] Taking this to the extreme would result in arguing that all individuals should perform all their economic activities in absolute independence from each other and that they should coordinate all these activities through market and price mechanisms. In more practical and modern terms, this would represent a "world of freelancers".

Unquestionably, this is not what can be observed in the real world. Besides individual economic players, the real world consists of firms, trade unions and many other kinds of "organizations" as understood herein (see above). One of the main reasons for the existence of such organizations instead of a completely free market of self-employed individuals has first been examined by Ronald Coase (1937) in his seminal article on *"The Nature of the Firm"*. In this article, Coase asks – assuming that market and especially price mechanisms are the best way to allocate resources for production and thus to coordinate economic activities – why organizations and large firms that supersede the price mechanism by organizational planning do actually evolve at all.

Coase identifies the *"cost of using the price mechanism"* as the main reason for the formation of organizations. These costs include expenditures for discovering relevant prices at the market, *"the costs for negotiating and concluding a separate contract for each exchange transaction which takes place on a market"* and costs that arise from the incompleteness especially of long-term-contracts. Today, these different kinds of costs are usually subsumed under the term *"transaction costs"*.[8]

[7] Mankiw and Taylor (2006, pp. 50 ff), for example, use numerical examples with gardeners and farmers or even with Robinson Crusoe and Man Friday to explain the principle of comparative advantage.

[8] There is a wide variety of definitions for transaction costs. One of the earliest mentions of the term refers to them as *"the costs of running the economic system"* (Arrow 1969, p. 48, as quoted by Williamson 1985, p. 8). Mankiw and Taylor (2006, p. 197) define transaction costs as *"the costs*

Different from the original "Coasian" understanding, such transaction costs can be argued to basically arise for *every* transaction being performed between two or more players – independently from the organizational framework under which the transaction is conducted. Williamson, for example, always considers transaction costs in the Coasian sense and what will be called the "costs of organizedness" herein as transaction costs conjointly and studies their characteristics for different organizational models (market, hierarchy, and hybrid). A similar concept of transaction costs is also applied by Milgrom and Roberts (1992). To avoid misunderstandings, we will thus refer to the "Coasian" transaction costs – the costs of using the market mechanism – as "market costs" from here on while the costs arising from transactions being coordinated through hierarchical structures will be considered as "costs of organizedness".

As mentioned above, any mechanism for fostering cooperation of different players has to solve the two main tasks of coordination and motivation. With cooperation being realized through the market mechanism, the task of coordination is simply solved via the price mechanism. For the price mechanism to be effective in coordinating individual activities and allocating other resources, the above-mentioned market costs have to be borne by the involved parties: Exchange partners have to be found, negotiations have to be made and contracts have to be monitored.

The task of motivation is, for the market-based case, usually assumed to be solved by individual self-interest of the involved players alone. Once the task of coordination is solved through the price mechanism, the simple principle of the "invisible hand" makes any further motivation dispensable.[9] Nonetheless, it can be necessary to spend further motivation costs under certain circumstances of market-mediated cooperation, too. These costs arise from *"costly arrangements [having to] be made to protect against opportunistic behavior"* resulting from *"information incompleteness and asymmetries"* and *"imperfect commitment"* between the involved parties of a transaction (Milgrom and Roberts 1992, pp. 29 f).[10] Most of these problems also apply for the case of hierarchical cooperation and will be examined in more detail in section 3.2.2. Using the price mechanism thus induces market costs for all involved parties and in return affords coordination and motivation simultaneously, whereas further motivation costs can arise under certain circumstances (see figure 3.2).

Due to the existence of these two types of costs, it can be highly expensive to coordinate individual activities on the pure basis of market mechanisms. Hierarchical structures may have an economic advantage over market mechanisms here, as they are able to supersede the price mechanism and thereby eliminate or at least reduce

---

that parties incur in the process of agreeing and following through on a bargain", while Eggertsson (1990, p. 14) defines them as *"the costs that arise when individuals exchange ownership rights to economic assets and enforce their exclusive rights"* and Williamson (1985, p. 19) simply describes transaction costs as *"the economic equivalent of friction in physical systems"*.

[9] See, for example, Milgrom and Roberts (1992, p. 28): *"People do not have to be cajoled, artificially induced, or forced to do their parts in a well-functioning market system."*

[10] Typically, such costly arrangements are more necessary with higher asset specificity, with uncertainty and complexity being present, where measurement of performance is difficult and where different transactions are connected with each other. See Milgrom and Roberts (1992, pp. 30 ff) or Williamson (1985, pp. 52 ff).

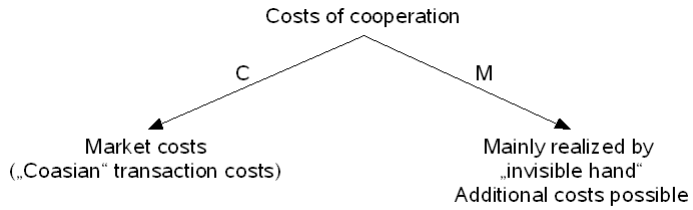


Figure 3.2: Coordination (C) and motivation (M) costs in the market model

market costs.[11] With significant market costs being present for certain transactions, hierarchical structures are thus able to govern activities more efficiently than the market could. It is this ability of "eliminating" market costs that is nowadays widely accepted as the economic explanation for the existence of organizations instead of purely market-driven cooperation.

But if organizations were only minimizing market costs and thus always had an advantage over market-based transactions, one had to ask immediately why *"not all production [is] carried on by one big firm"* (Coase 1937). The explanation is that organizations also face *"costs of organizedness"*, which can reach considerable amounts, too. These are addressed in section 3.2.

Conclusion:
Cooperation of individual players through the price mechanism entails market costs for solving the task of coordination. The price mechanism also solves the task of motivation to a certain extent, while further motivation costs can arise under specific circumstances. The existence of these two types of costs is the main reason for the formation of organizations instead of using mechanisms of a completely free market to realize cooperation.

3.2 Costs of Organizedness

Organizations face, in contrast to the idealized model of a free market, other kinds of costs which represent the limiting factor for their growth. Coase (1937) not only recognized market costs as the constitutive reason for the existence of firms (and other organizations), but also identified the costs arising from performing transactions *inside* organizations as determining factor for the size of these organizations.

[11] See, for example, Coase (1937): *"It can [...] be assumed that the distinguishing mark of the firm is the supersession of the price mechanism."* See also Coase (2005, p. 34): *"[The] existence [of market costs] implies that methods of coordination alternative to the market, which are themselves costly and in various ways imperfect, may nonetheless be preferable to relying on the pricing mechanism."* Furubotn and Richter (2005, p. 369) ascribe the cost advantage not to the absolute supersession of the price mechanism but rather to a significantly lower number of contracts having to be closed in the hierarchical case.

Again, these "costs of organizedness" can be subdivided into two major areas: coordination costs and motivation costs. For the case of organizations, coordination costs mainly arise from efficiency losses which will be considered in more detail in section 3.2.1 while motivation costs can mainly be assigned to the term of "agency costs" which are addressed in section 3.2.2.

3.2.1 Hierarchical Coordination Costs

The first part of the "costs of organizedness" refers to losses resulting from the organization's inability to allocate resources optimally. As Coase (1937) points out for the example of firms, *"[a]s the transactions which are organized increase, the entrepreneur fails to place the factors of production in the uses where their value is greatest [...]."* These efficiency losses are widely known in economics as *diminishing returns to management* and of course also exist for other kinds of organizations than firms. With increasing size (in the sense of an increasing number of transactions being realized internally), organizations suffer substantial disadvantages in opposition to the model of free market exchange. Obviously, this effect works against the size-increasing effect of organizations as "market costs eliminators" mentioned above.

The inner principles leading to the mentioned diminishing returns to management have for long times been subject to economic discussions.[12] Knight (1921, p. 286) supposes an increasing amount of uncertainty under which decisions have – due to limitations *"of the scope of enterprise one man can deal with effectively"* – to be made in a hierarchical structure as explanation. Williamson (1985, p. 133) gives the counterargument that *"decisions need not be forced to the top but can always be assigned to the level at which the issues are most appropriately resolved."* By letting decisions be made at lower hierarchical levels, the capacity-problem of decision-makers could thus be overcome. But at the same time, increasing the organization's size would in this model *"necessarily [entail] adding hierarchical levels"* (id., p. 134). As communication between each of these levels always leads to observable losses and alterations of communicated content and thus entails what Williamson calls *"control loss"* (id.), every additional hierarchical level leads to an additional loss of control within the organization. In this model, the efficiency losses within organizations and the resulting limitations of organizational size are thus ascribed to the fact that the cumulative "control loss" increases exponentially with an increasing number of existing hierarchy levels and *"eventually exceed[s] the gains"* (id.) from organizational elimination of transaction costs.[13]

However, Williamson himself puts this model into question as being the *sole* explanation for the limited size of organizations. Again, making selected decisions at lower hierarchy-levels could also diminish the amount of *control loss* to a certain ex-

[12] For a short overview on different attempts of explanation, see Williamson (1985, pp. 132 ff).

[13] The explanation given by Milgrom and Roberts (1992, p. 29) points into the same direction. They ascribe the diminishing returns to management to the costs of communicating information through the hierarchy and of determining an efficient plan from the gathered information. Milgrom and Roberts especially mention decision-makers always having *"only insufficient or inaccurate information"* what ultimately leads to *"maladaption"*.

tent. Even if being more explanatory than the model given by Knight (1921), the model does thus not explain why organizations do not make use of the mentioned kind of *"selective intervention"* to overcome the limitations of organizedness. As an alternative explanation for the diminishing returns to management, Williamson thus develops the model of *"the costs of bureaucracy"* (1985, pp. 148 ff). These costs are assumed to arise mainly because of decision-makers having a *"propensity to manage"* – a well-intentioned but failing *"instrumental"* as well as a selfish *"strategic propensity"*[14] – and because of *"forgiveness"* being prevalent in organizations as a result of different transactions being interdependent.[15] The resulting costs of bureaucracy are assumed to increase with the size of the organization and thus to represent – at least – another limiting factor for the size of organizations.

Selecting one of these explanatory approaches as the sole source of diminishing returns to management inside organizations is not necessary for the scope of this work. In all likelihood, every phenomenon will have its right to exist and allows for valuable insights. As basic principles inducing coordination costs inside hierarchical organizations, we can thus take into account

- a central managements inability to gather and incorporate *all* relevant information into decision making[16] resulting in a misallocation of resources (efficiency losses) and
- a general propensity to "overmanage" conduct inside organizations and a tendency to make decisions not only on the basis of economic but also of political and strategic considerations (bureaucracy costs).

Furthermore, these different sources of coordination costs inside organizations can also interact with each other. For example, it could be decided on the basis of strategic and political considerations that a certain kind of conduct should be managed centrally and without selective intervention at lower levels. The result of the strategic and political considerations would then, due to more communicative losses than necessary, be a higher misallocation of resources.

All these costs refer to the problem of efficient allocation of activities and other resources inside organizations and are thus part of the *coordination costs*. The corresponding *motivation costs* arising for the cooperation model of the hierarchical organization will be discussed in the following section.

[14] In both cases, the propensity to manage leads to managers trying to manage more subjects than it would be optimal. In the *"instrumental"* case, this happens because of managers overestimating the own capabilities while in the *"strategic"* case, managers consciously accept inefficiencies for selfish reasons. For this *"strategic propensity"*, see also Powell (1987, p. 79): *"There is a pronounced tendency in large organizations for managers to cause their firms to grown beyond the optimal size. More subordinates apparently satisfies a manager's psychological needs, increases his or her social status within the firm, and provides tangible economic rewards [...]"*.

[15] Furthermore, Williamson (1985) also mentions the fact of *"internal operating and investment decisions [being] [...] subject to politicization"* instead of economic considerations as a further source of bureaucracy costs (pp. 151 f).

[16] Alternatively, one could also refer to the enormous costs that would arise from any attempt to gather and take into account the respective information. The effect would basically be the same.

3.2.2 Hierarchical Motivation Costs

Besides the costs arising from the inability to determine an optimal allocation of productive resources, organizations also face costs resulting from members not being motivated to act in the best interest of the organization. These costs are usually subsumed under the term of *"agency costs"*, which was widely established by Jensen and Meckling (1976).[17] Even if the basic relation considered by Jensen and Meckling is the one between owners and the top management of a corporation, their definition of an agency relationship as *"a contract under which one or more persons (the principal(s)) engage another person (the agent) to perform some service on their behalf which involves delegating some decision making authority to the agent"* (id., sect. 1.4) goes far beyond this very specific case and also includes other hierarchical relations.[18]

The fundamental economic insight from principal-agent theory considering such kinds of relations between two parties is that *"[i]f both parties [...] are utility maximizers, there is good reason to believe that the agent will not always act in the best interests of the principal"* (Jensen and Meckling 1976, sect. 1.4) because the utility functions of the agent and the principal usually differ for the different possible outcomes of a specific decision having to be made by the agent (Eggertsson 1990, p. 41). In such situations, the above-mentioned agency costs arise because of two basic principles which will be examined below: information asymmetries between principal and agent and the possibility of opportunistic behavior and moral hazard by the agent as a result of the divergent utility functions in connection with information asymmetries. The opportunistic behavior of the agent then causes suboptimal outcomes for the organization and thus produces costs which are clearly subject to the task of motivation having to be solved. To reduce these costs, there exist different countermeasures which reduce opportunistic behavior (or: motivate the agent to act more in the interest of the principal) but at the same time induce other kinds of costs.

3.2.2.1 Information Asymmetries

Information asymmetries are the main reason for inefficiencies within principal-agent relations. Basically, they can be defined as *"difference[s] in access to relevant knowledge"* (Mankiw and Taylor 2006, p. 446) and they thus arise whenever two parties interact with each other and have different knowledge about the conditions under which the interaction takes place. Three types of information asymmetries can be distinguished: *Hidden characteristics* refer to the "quality" of an agent (his capabilities or his efficiency, for example) or an exchanged good, *hidden action* to activities of one involved party that cannot be completely observed by the other and *hidden information* stands for situations where one party has more knowledge about the givens of the situation a certain activity is conducted in. Such information asymmetries play an outstanding role for numerous areas of modern economics and their existence is one

[17] For an introductory and easy to understand overview of agency theory, see also Sappington (1991).

[18] In fact, Jensen and Meckling (1976) also had these relations in mind: *"[The problem of principal-agent relations] exists in all organizations and in all cooperative efforts—at every level of management in firms, in universities, in mutual companies, [...]"* (sect. 1.4, footnote omitted).

of its constitutive fundamentals.[19]

In principal-agent relations, the problems of hidden action and hidden information are most decisive:[20] The agent is engaged to perform an activity on behalf of the principal and is given a certain amount of autonomy for this purpose.[21] This autonomy necessarily involves the agent to have a certain amount of information that is not known by the principal – be it detailed knowledge about the specific conditions under which he has to conduct the activity or the effort that the agent makes. There is thus an information asymmetry about the activity performed by the agent, allowing him to conduct opportunistic behavior and thereby arising the problem of moral hazard.

3.2.2.2 Opportunism and Moral Hazard

As the agent will be conscious of the principal's limited information about his actual behavior, he will have an individual interest in investing less effort than would be optimal from the viewpoint of the principal: Due to the above-mentioned difference of the agent's and the principal's utility functions, the opportunistic agent would prefer shirking over diligence – at least as long as he is paid a fixed salary and as no further incentives for investing more efforts are present.

To bar the agent from such opportunistic shirking, the principal has the first option to limit the agent's autonomy and thereby prevent opportunistic decisions. This would decrease the costs arising from selfish behavior, but would in turn remove any benefits possibly arising from delegation of decision-making and thus increase efficiency losses (see above).

Second, the principal can try to evaluate the agent's actual effort and performance on the basis of some kind of monitoring. The agent could then be motivated to act in the interest of the principal – that is, to behave in the way that has been "assigned" to him in the previous process of coordination – either by enforcing the respective behavior or through rewarding obedience. In both cases, the agent gets an incentive to behave in the way that is intended by the principal and the agent's interests are thus aligned with those of the principal – at least to a certain extent.[22] Even if all-encompassing monitoring would contradict the principle of delegation of activities that characterizes principal-agent relations, a certain extent of monitoring might induce less *monitoring costs* than it provides returns to the principal (through motivating the agent to invest more efforts). Nonetheless, monitoring only pays off to a certain extent and a certain amount of opportunistic behavior will always remain.[23]

[19] Akerlof (1970), for example, explained the principle of adverse selection and the possibility of a resulting market failure on the basis of hidden information in the market for used cars.

[20] Nonetheless, hidden characteristics like the talent of an employee might also play an important role. For principal-agent relations, these hidden characteristics are primarily of relevance *ex-ante*, before a specific agent is chosen. See, for example, Spence (1973).

[21] Remember the idea of "selective intervention" mentioned above.

[22] For an extensive and highly formal treatise on incentives in principal-agent relations, see especially Laffont and Martimort (2001).

[23] See Eggertsson (1990, pp. 42 f): *"The marginal rate of return on resources invested to constrain the agents fall after a point, and in most cases it does not pay to try to eliminate all opportunistic behavior."*

Furthermore, monitoring can in many cases only be conducted efficiently by measuring some secondary indicators[24] instead of observing the agent's behavior directly.[25] These indicators could be the time spent in the office, the number of contracts having been closed with customers or even the number of published academic papers as it is now usual practice in many universities to evaluate scientific personnel.

As such secondary indicators do not perfectly represent the principal's original objectives, the agent can concentrate on serving these imperfect indicators instead of behaving the way that was originally intended by the principal. The agent is thus not motivated to behave exactly in the way that had been determined as being optimal during the process of coordinating activities (see above). The same is also true for the use of other productive resources, where the agent might face incentives to use resources opportunistically for his own goals because the use of the resources cannot be monitored perfectly. Additionally, the agent can also exploit the principal's limited knowledge about the circumstances under which a certain activity had to be conducted by ascribing bad outcomes to disadvantageous exogenous events while attributing good outcomes to large efforts having been made by himself.[26] This problem is usually referred to as *"moral hazard"*.

And finally, a third possible approach for solving the motivation problem in principal-agent relations shall, for the sake of completeness, also be mentioned in brief: the use of *bonds*, which the agent deposits with the principal and which are *"forfeited if the agent is caught cheating"* (Milgrom and Roberts 1992, p. 195).[27] Like the other mechanisms mentioned above, bonds are able to align the agent's interests with those of the principal and can limit the costs arising from opportunistic behavior of the agent. Again, a certain amount of (bonding) costs has to be borne to motivate the agent to behave the way that was planned by the principal during the coordination process.

Altogether, cooperation in the hierarchical model thus basically leads to coordination costs in the form of efficiency losses and bureaucracy costs. For solving the task of motivation, additional agency costs have to be borne (see figure 3.3).

[24] Sometimes these indicators are also called "proxies". See, for example, Voigt (2002, p. 102).

[25] In some cases the use of secondary indicators is not even subject to efficiency considerations but rather necessary as a matter of principle. See, for example Milgrom and Roberts (1992, p. 170): *"[T]he employee's supervisor cannot determine whether he or she is thinking about company business or personal matters."*

[26] See, for instance, Voigt (2002, p. 104) or Furubotn and Richter (2005, p. 163).

[27] This mechanism and the respective *"bonding costs"* have also been considered by Jensen and Meckling (1976, sect. 1.4): *"In addition in some situations it will pay the agent to expend resources (bonding costs) to guarantee that he will not take certain actions which would harm the principal or to ensure that the principal will be compensated if he does take such actions."*

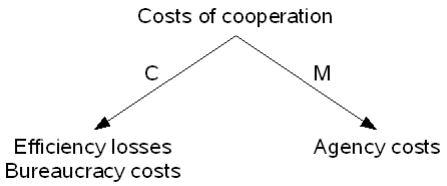


Figure 3.3: Coordination (C) and motivation (M) costs in the hierarchical model

Conclusion:
Even if hierarchical organizations are able to eliminate substantial amounts of "market costs" and thus can have advantages over market-based cooperation, they have to bear other kinds of costs. The task of coordination being conducted in hierarchical organizations leads to substantial efficiency losses due to a central management's inability to gather and incorporate all relevant information in decision making and because of different factors resulting in "costs of bureaucracy". Solving the task of motivation inside hierarchical organizations entails further costs for monitoring the agent, for depositing bonds and because of a certain amount of remaining opportunistic agent behavior. The alternative of restricting the agent's autonomy would decrease motivation costs but on the other hand increase efficiency losses.

3.3 Market Costs, Costs of Organizedness and Hybrid Models

We have so far shown that individuals can benefit from cooperating with each other because of the possibilities for labor division and specialization. Nonetheless, this cooperation has to be realized under a framework that allows to *coordinate* individual activities and the efficient allocation of other factors and that *motivates* individuals to actually behave in the way that was deemed advantageous during the coordination process. Solving these tasks is not possible for free but always involves further costs, which can be subsumed under a wide definition of *transaction costs* (Williamson 1985). The distinction of *coordination costs* and *motivation costs* as developed by Milgrom and Roberts (1992, pp. 29 f) allows for more structured considerations about the different kinds of costs arising in different models of cooperation.

Two idealized frameworks for conducting cooperation can be identified: free market exchange and the hierarchical organization. These ideal types face different kinds of costs having to be borne to solve the tasks of coordination and motivation. With market-based cooperation, coordination of activities and factor allocation is solely left to the "invisible hand" of the price mechanism that also provides motivating incentives to the involved parties. Using this price mechanism entails "market costs" as described in section 3.1. Nonetheless, even with market-based cooperation, there might be situations where principal-agent relations evolve and entail further agency costs (see

section 3.2.2). With cooperation being conducted within hierarchical organizations, the task of coordination produces costs in the form of control loss and a wide variety of bureaucracy costs as outlined in section 3.2.1 and the task of motivation again gives rise to agency costs (3.2.2). To avoid misinterpretations, it has to be noted that the agency costs occurring under market-based cooperation are not congruent with those emerging in hierarchy-mediated cooperation, for example because of the price mechanism providing inherent incentives to the involved parties. In a highly simplified model, we can thus outline the different cost structures of market- and hierarchy-based cooperation as done in figure 3.4.

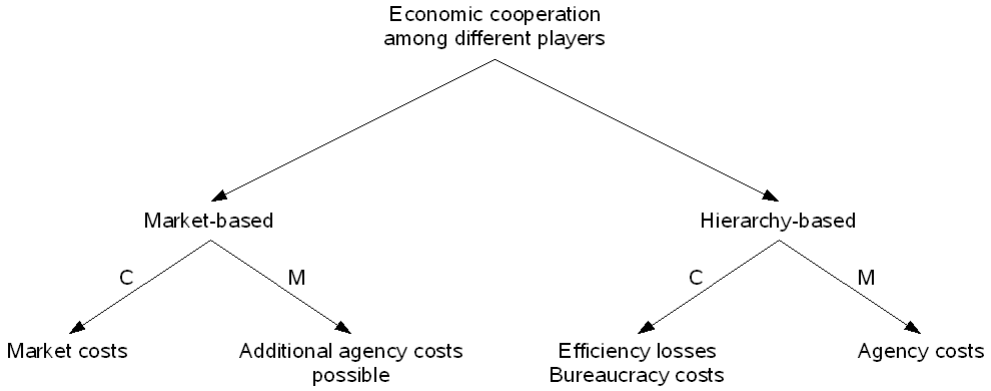


Figure 3.4: Coordination (C) and motivation (M) costs for market- and hierarchy-based cooperation (simplified model).

Leaving further factors aside for demonstrative purposes, the decision under which kind of cooperation a specific type of transaction should be conducted can be made on the basis of the different respective cost structures resulting from market- and hierarchy-based cooperation. Generally speaking, the market-based alternative should be chosen when market costs are low – because of the low specificity of a commodity good to be transferred, for example – and where the price mechanism provides enough incentives. The hierarchy-based modus should in turn be chosen where extensive market costs would – due to high specificity, for example – outreach the costs having to be borne because of hierarchy-internal control-loss, bureaucracy costs and further motivational arrangements having to be made.

However, only seldom will one of the idealized models of cooperation through pure markets or pure hierarchies be the most beneficial one. Instead, cooperation between different players is in the majority of cases realized through a mix of practices originating from the world of markets as well as from the world of hierarchy. Williamson (1991) terms such arrangements *"hybrids"*, which are *"characterized by semi-strong incentives [and] an intermediate degree of administrative apparatus [...]"* (p. 281). These hybrids can exist in the form of franchising, alliances, partnerships or cooperatives, for example.[28]

[28]See Ménard (2005, p. 295). For further examples, see also Furubotn and Richter (2005, p. 386) or

As Hennart (1993) points out, most transactions are in the real world performed under such a hybrid mode of cooperation – with a continuum ranging from nearly perfect markets to nearly perfect hierarchies and with a *"swollen middle"* of cooperation being performed near to the medium in most cases.[29] Hennart shows that this *"swollen middle"* emerges as a result of cost considerations: Because of market costs as well as costs of organizedness increasing strongly at the respective extremes, minimization of the overall sum (the overall costs of cooperation) leads to intermediate outcomes and thus to cooperation being realized through hybrid arrangements.

For our purposes, we can thus distinguish three general modes for realizing economic cooperation amongst different players: markets, hierarchies and different kinds of hybrids lying between those two extremes. All of these different modes solve the tasks of coordination and motivation in a different way and feature different cost structures. Market-like cooperation has cost advantages especially for transactions of low specificity and complexity where the task of motivation can be solved by the *"invisible hand"* of the price mechanism alone. Hierarchical cooperation is, in turn, generally advantageous when the transaction is highly specific and complex and where cooperation through market-like mechanisms would thus entail substantive additional costs for solving the task of motivation. Between those two extremes lies a wide variety of hybrid arrangements that prove to be economically advantageous for most cases because of providing an optimized relation of *"market costs"* and *"costs of organizedness"*.

The fact of such a continuum of possible forms of cooperation existing between the extremes of "market" and "hierarchy" was also demonstrated in a vivid manner by Malone (2004), who delineates a *"decentralization continuum"* ranging from centralized hierarchies over loose hierarchies and democracies to markets (see figure 3.5).

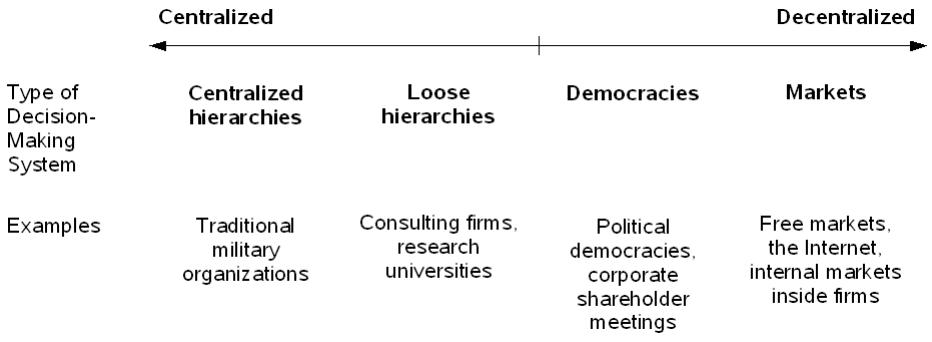


Figure 3.5: "Decentralization Continuum" according to Malone (2004, p. 6)

Klein (2005, pp. 448 f). For a slightly more legal perspective on hybrid arrangements, see Geis (2009).

[29] The fact of a continuum being present between the extremes is also highlighted by Furubotn and Richter (2005, p. 291): *"In principle, the institutional solution adopted can be one of the two extreme forms (markets or hierarchies) or anything in between."* See also Klein (2005, p. 438): *"Organizational form is often modeled as a discrete variable—'make,' 'buy,' or 'hybrid,' for example—though it can sometimes be represented by a continuous variable."*

Less bound to the economic apparatus of transaction costs outlined above, Malone primarily ascribes the choice of a specific level of centralization from this continuum to the *communication costs* having to be borne. According to Malone, the lower these communication costs are, the more decentralized will the applied model of cooperation be. And with new technologies constantly lowering these communication costs, ongoing technological change results in more and more cooperation being organized in a decentralized manner. This relation between technology and the chosen model of cooperation was not only identified by Malone and as we will see, it fits perfectly into the model of cost structures developed above. One final aspect will thus complete this chapter: the effect of technology and especially of technological change on the cost structures of different mechanisms for realizing cooperation.

Conclusion:
The idealized models of cooperation either being realized through market mechanisms or through strictly hierarchical structures have to be understood as extreme cases which represent the optimal mode for very few cases. From the perspective of cooperation costs having to be borne, hybrid modes of cooperation lying somewhere between those extremes allow to reduce overall costs and thus prove to be advantageous for most cases.

3.4 Organizational Models and Impact of Technology

The impact of technology and technological change on transaction costs and thereby on the resulting organizational structures was already recognized by Coase in his article on "the nature of the firm" mentioned above. Coase considered the at that time increasing use of telephones and stated that *"[c]hanges like the telephone [...] which tend to reduce the cost of organizing spatially will tend to increase the size of the firm."* But at the same time, Coase was also aware that *"most inventions will change both the costs of organizing and the costs of using the price mechanism"* and that the actual effect of technological change will thus *"depend on the relative effect on these two sets of costs."* (Coase 1937, p. 397)

Consistent with Coase's original perception, Malone, Yates, and Benjamin (1987, pp. 487f) explain how the development and diffusion of the telegraph led – by making hierarchical cooperation over wide distances possible for the first time – to larger organizations establishing in a first stage. Importantly, this did not happen by substituting market-based cooperation but by allowing formerly non-existing cooperation with remote partners at all. In terms of costs as outlined above, costs of cooperation were thus prohibitively high before the telegraph, thereby preventing *any* kind of remote cooperation, be it under a market-like or under a hierarchical regime. The telegraph then changed these conditions and led to cost structures under which the benefits from hierarchical cooperation exceeded the respective costs of organizedness, giving rise to larger and locally dispersed hierarchical organizations.

Regarding more recent developments in the field of information technology, Malone

et al. (1987, pp. 488 f) identify an *"electronic brokerage effect"* that allows for electronic markets to be realized and that leads to significantly lower costs of conducting exchange (and thus, cooperation) through more market-like structures. Within our model of cooperation costs, this effect can be described as lowering the coordination part of market costs by significantly reducing search costs. Even if Malone et al. (p. 489) also admit that *"the effects of information technology [...] make both markets and hierarchies more efficient,"* they conclude that *"information technology will lead to an overall shift toward proportionately more use of markets rather than hierarchies to coordinate economic activity"* (p. 496).[30] Later on, Malone (2004) derived his *"amazing pattern"* from these considerations: Starting from small independent businesses, the prevailing model of cooperation changed to large hierarchies because of dropping communication costs. With these communication costs decreasing even further, hierarchical models are then more and more transformed to rather market like structures (see figure 3.6).

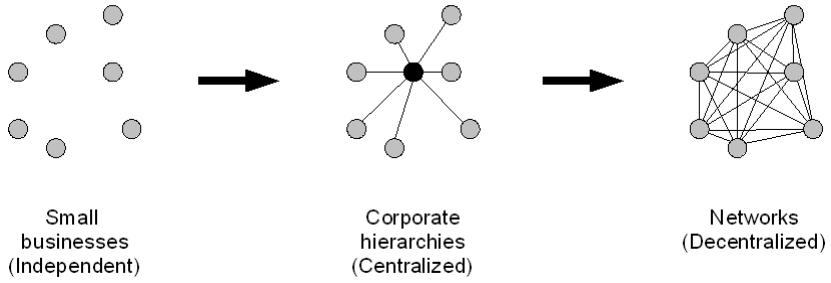


Figure 3.6: "Amazing Pattern" according to Malone (2004, p. 28)

More linked to the transaction cost approach of Coase and Williamson used herein, the argument of information technology mainly lowering the costs for market-like cooperation was also brought forward by Ciborra (1985, p. 63), who characterizes information technology as *"a means for creating/expanding markets, by lowering search, contracting and control costs."* Nooteboom (1992, pp. 343 f) points into the same direction by highlighting the effect of information technology increasing the *"transparency of supply and demand"*. Through this increase, search costs can be minimized and more transactions should be realized through market mechanisms.

On the other hand, information technology obviously does not only lower the costs for market-like cooperation but can also have decreasing effects for the costs arising from hierarchical cooperation.[31] For example, information technology can lower the efficiency losses mentioned in section 3.2.1 through the use of highly integrated management information systems. Such systems can provide managers with more and

[30] For a graphical and more formal explanation, see also Brynjolfsson, Malone, and Gurbaxani (1988, pp. 5 ff). Note that the term of *"coordination"* is used in a slightly different manner by Malone et al. than it is herein, including all aspects of cooperation. Nonetheless, this distinction is of minor importance for the general point made here.

[31] For empirical data supporting this argument, see, for example, Shin (1997).

better information regarding the decision having to be made and thereby mitigate the *"capacity problem"* of managers to a certain extent. Similarly, information technology can also help to overcome the problem of *"control loss"* by reducing loss and alterations of information that is communicated through different hierarchy levels.[32] Besides these two aspects of coordination costs, information technology could also help reducing motivation costs for the hierarchical case. Monitoring costs, for example, can be significantly reduced through information technology, thereby alleviating information asymmetries between principal and agent and thus lowering the agency costs arising in the hierarchical model.[33] Information technology can thus lower the costs of cooperation for the hierarchical case, too.

As it becomes clear, the impact of technological change to the chosen model of cooperation can be multifold. Electronic markets can especially reduce search costs, thereby limiting market costs and causing a shift toward more market-like arrangements for conducting cooperation. Other technologies might allow the management of a hierarchical organization to incorporate more information into decision making or to monitor employees more efficiently. Such changes would represent a decline of the efficiency losses and agency costs mentioned in section 3.2, result in lower costs of organizedness and thus lead to a shift toward more hierarchical modes of cooperation.

Of those two directions, literature suggests the effect of reducing market costs to be the more significant one in the majority of cases. Brynjolfsson et al. (1994), for example, found strong indications for the size of organizations to decrease with increasing use of information technology. Nooteboom (1992, p. 349) makes a comparable argument when stating that, *"[o]n the whole, ICT tends to reduce transaction costs and thereby confirms the tendency to buy rather than make"*. Dedrick and Kraemer (2005) analyzed the personal computer industry in detail and show that *"most PC makers have narrowed their scope of activities"* (p. 128) significantly – by outsourcing the manufacturing of components, for example – and increasingly concentrated on strategic and complex issues like build-to-order processes, marketing or customer-service. The Internet as well as electronic data interchange between PC makers and their different suppliers played a significant role in this change toward using more market mechanisms.[34] Earle, Pagano, and Lesi (2002) detect information technology to have had

[32] See, for example Gurbaxani and Whang (1991, p. 69): *"IT enables organizations to process decision-relevant information in a more cost-effective way, thus improving the quality and speed of upper management's decisionmaking processes [...], leading to more centralized management."* For an example of how information technology can help to handle the same activities with fewer people in a hierarchical model, see Pinsonneault and Kraemer (2002, pp. 198 ff). See also Dewett and Jones (2001, pp. 329 f).

[33] See again Gurbaxani and Whang (1991, p. 69): *"IT can also provide management with the ability to reduce agency costs through improved monitoring capabilities and performance evaluation schemes"*. See also Brynjolfsson, Malone, Gurbaxani, and Kambil (1994, p. 1632) mentioning a reduction of agency costs as a possible outcome of information technology being used in hierarchical organizations.

[34] These findings notwithstanding, Dedrick and Kraemer interestingly state that they did *"not find an inherent bias toward market transactions as a result of IT or the Internet"* (p. 139). Instead, they point to *"more specific issues"* like *"complexity of products and processes"* – and thus, of transactions – being more significant. The fact that information technology usually lowers the effective complexity of transactions by making them more controllable seems to have been over-

a decreasing effect on firm sizes in the *"quasi-experiment"* of economic transition in Central and Eastern Europe. And finally, Malone (2004) also gives ample evidence for the decentralizing and size-decreasing effect of information technology to prevail over the advantages being gained by hierarchies.

Furthermore, Malone clarifies that hybrid modes of cooperation can, besides the hybrid-specific approaches mentioned by Williamson (1991, pp. 280 ff), also be realized through a *mix* of approaches being used simultaneously by one and the same group of cooperating individuals. In this case, some activities are conducted in a strictly hierarchical manner and others through organization-internal market mechanisms, for example. Mentioning typically less-hierarchical organizations like universities as well as large and usually hierarchical industrial organizations like electric power suppliers or fabric makers, Malone demonstrates how organizations can profit from a mix of centralized and decentralized processes: Opinion polling, organization-internal freelancing and different kinds of (quasi-) markets are examples for mechanisms that can – especially suppported by information technology – be used inside organizations to overcome the different shortcomings of classical hierarchies. A hybrid, cost optimizing arrangement for cooperation can thus also be realized by employing hierarchical practices for some cooperative activities and more market-like mechanisms for other ones.

In any case, it has to be noticed that information technology typically changes costs structures in both aspects – hierarchy costs as well as market costs – and allows for non-hierarchical modes of cooperation to become economically advantageous over the hierarchical model in certain cases. In particular, information technology can help establishing decentralized mechanisms or even render them possible at all. In most cases, however, *"the best solution is to create a custom system that combines elements of more than one basic structure"* (Malone 2004, p. 113).

Conclusion:
Technology and technological change significantly affect cost structures of different modes of cooperation. Hierarchy costs as well as market costs can be reduced in a variety of ways through adoption of technologies like management information systems, monitoring technologies, electronic markets or electronic reduction of the effective complexity of transactions. Of those two effects, literature suggest the market-directed effect to outreach the advantages for hierarchical models, thereby leading to an overall shift toward more cooperation being realized under less hierarchical regimes. Furthermore, technology and technological change also allow hybrid modes of cooperation that use hierarchical and market-like mechanisms conjointly.

looked by the authors – even if they themselves state that *"The Internet [...] has reduced asset specifity and reduced the costs associated with market transactions"* (p. 124). Nonetheless, the decentralizing effect can only appear if information technology is used in a "complexity-reducing" manner, of course. Malone (2004, p. 31) also mentions the computer industry as example for more cooperation being realized in a non-hierarchical manner.

3.5 Conclusion

This chapter addressed essential economic aspects of organizations to allow for well-founded considerations of information security *inside* organizations in subsequent parts. As we have shown, the existence of organizations can be ascribed to the basic principle of labor division and specialization leading to better overall outcomes. To realize those better outcomes, different individuals with different specializations have to cooperate, whereas any mechanism for realizing this cooperation has to solve two general tasks: First, it has to assign different activities to the different people so that every individual engages in those activities leading to the largest overall benefit. This task can be referred to as the *task of coordination*. And second, any mechanism also has to motivate the respective individuals to really engage in those activities assigned to them during the coordination process. Besides the task of coordination, there is thus also a *task of motivation* having to be solved by any mechanism for realizing cooperation.

Two generalized models can be identified for solving those tasks: markets and hierarchies. Both are able to coordinate activities and to motivate individuals to behave in the intended manner, but both also give rise to additional costs having to be borne. These *costs of cooperation* differ significantly between market- and hierarchy-based cooperation. Using market mechanisms for the task of coordination leads to *market costs* as first described by Coase (1937): potential exchange partners have to be found, prices have to be negotiated, contracts have to be drawn, etc. In the idealized market, motivation is solely realized by the *invisible hand* of individual self-interest, but in many cases, additional costs have to be borne because of existing information asymmetries and the resulting agency problem, for example.

Due to these *"costs of using the price mechanism"* (Coase 1937), it can be advantageous to realize cooperation through the alternative model of hierarchies. These are able to eliminate search costs, bargaining costs, etc. but at the same time lead to other kinds of costs. *Coordination costs*, for example, arise from information being imperfectly communicated through hierarchy levels or from management's decisions not being solely based on economic aspects but also on political and strategic considerations. Furthermore, *motivation costs* also emerge in the hierarchical model and are – due to the absence of an *invisible hand* – typically more significant than in the market-based case. Both models thus entail different kinds of costs, leading to the challenge of identifying the model of cooperation that minimizes overall costs for a given case. The underlying decision between *"make or buy"*, between cooperating through hierarchies or through markets is one of the classical problems of economics.

Established findings suggest that the cost optimizing model will not be one of the idealized extreme forms of market or hierarchy, but rather a hybrid mode of cooperation for the majority of cases. Such hybrid modes can be located anywhere on a continuum between markets and hierarchies and can be realized through long-term contracts, franchising models, alliances or partnerships, for example. Furthermore, there is also the possibility of using different kinds of mechanisms within one and the same organization as well. For example, an otherwise strongly hierarchical organization can organize single activities through organization-internal markets and thereby

profit from the best of both worlds.

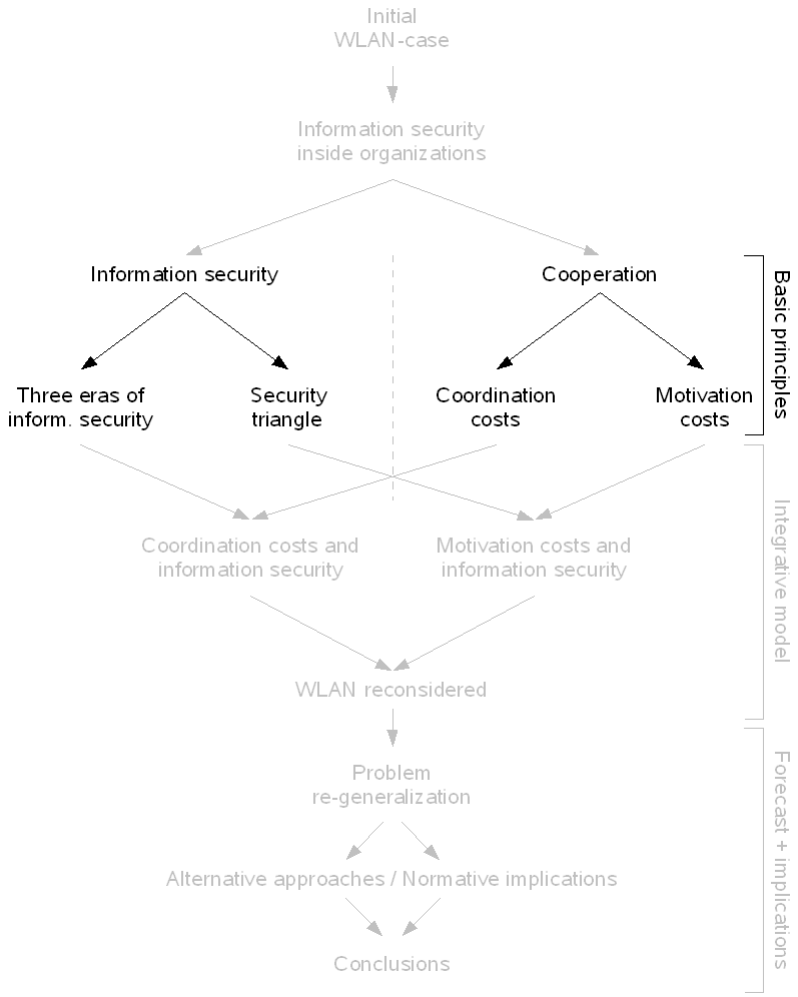
As a final aspect, we considered the impact of technology and technological change. As shown, technological change usually results in a change of cost structures. Information technology can have an impact on hierarchy costs – by lowering monitoring costs, for example – as well as it can affect market costs significantly. Progress in the field of information technology thus unquestionably affects the different cost structures of market- and hierarchy-based cooperation and will have to result in a different model being chosen for cooperation. Even if there have been certain disagreements whether this shift will be directed toward more market- or more hierarchy-based cooperation, current findings seem to identify the effect of reduced market costs to be the more significant one, leading to a growing part of cooperative activities being realized through non-hierarchical mechanisms.

With this chapter, we thus established the second pillar for well-founded considerations and for our positive model of *information security inside firms* to rest on. The distinction of the tasks of coordination and motivation and of the respective costs having to be borne will be of high importance for subsequent discussions. Similarly, the concept of hybrid arrangements and their possible expedience will also be taken on in later chapters.

As a first step toward bringing together information security and the economic perspective, the following chapter will outline some selected economic views on information security that have been developed over the past few years.

Chapter 4

Economic Perspectives on Information Security



Chapter 4

Economic Perspectives on Information Security

The renewal of a scientific discipline usually does not originate at the center [...].

– Thrainn Eggertsson

So far, the two fundamental areas for *information security inside organizations* – the status quo of information security and the economic foundations of organizations – have been examined separately. But to allow for well-founded considerations of information security inside organizations, these two areas have to be brought together. Over the past few years, first steps into this direction have been made in the discipline of *Information Security Economics*. Especially the yearly Workshops on the Economics of Information Security (WEIS[1]) have played a significant role for this development. As ongoing research has shown so far, economic principles can be applied to diverse problems arising in the field of information security:

Lack of security in operating systems and other software products can, for example, be explained by the economic nature of software markets. As these markets are in many cases driven by network effects[2], vendors have strong incentives to bring their products to the market as soon as possible to benefit from first-mover-advantages and to reach the critical mass of users. Resulting from these incentives, it can be *"perfectly rational behaviour"* for a software vendor to ship unfinished and thus insecure products to its customers just to take a temporary advantage over its rivals (Anderson 2001, p. 2). Furthermore, this effect is also supported by information asymmetries about the security of a software product, resulting in a *"lemons market"* (Akerlof 1970) with secure but expensive software not being demanded by customers.[3] The same mechanism of information asymmetries can not only be observed in markets for standard software like office-suites or operating systems, but even in the market for security products like antivirus software, firewalls or intrusion detection systems.[4]

[1] Proceedings and other information on the workshops can be found at `http://weis09.infosecon.net/past_workshops.html` [16.03.2009].

[2] For network effects and information economics, see especially Shapiro and Varian (1999).

[3] See again Anderson (2001, pp. 5 f).

[4] See, for example, Schneier (2007a): *"The few products that succeeded weren't the most secure, because buyers couldn't tell the difference."*

Again, this situation results from perfectly rational behavior of organizations – in the role of customers – and from software vendors acting perfectly rational, too.

But organizations have also been subject to economic considerations of information security more directly. Information Sharing and Analysis Centers (ISACs)[5], for example, are aimed at enhancing the information security for all of their member organizations through sharing of information about existing security breaches or successful solutions to known problems. Nonetheless, member organizations have to be considered as acting upon individual economic interests and might therefore be *"tempted to free-ride and under invest in new methods to deal with attempted and successful security breaches"* (Gordon et al. 2002). As Aviram (2006, pp. 152 ff) shows, this and other economic conditions lead to noncooperative game types being played within ISACs, necessitating additional economic incentives for individual member organizations to prevent the ISAC from failing. In this case, economic considerations are able to explain why ISACs fail and additionally suggest strategies for preventing them from doing so.

Economic principles have been applied to many more questions from the area of information security, but we will leave it at the mentioned two.[6] As those do, most considerations from the area of the economics of information security originate, in some manner, from a macro-perspective: When software vendors seem to have an interest in producing secure products and when not, how it can be achieved that different organizations share their knowledge about existing vulnerabilities, which amount an organization should invest in information security and against what kind of threats, etc.: In most cases, an organization is considered as a singular, uniform entity of the respective model, acting in a certain and well defined manner as a whole.

In fact, this does not adequately represent usual organizations. Rather, nearly every organization has to deal with individual job-profiles, interests, skills and habits of its members, with information that is hidden to central instances[7] and with a couple of additional conditions arising from the respective organizational structure. In short: Organizations are anything but uniform. They consist of people. And these people can be taken into account as individual economic actors for applying economic theory to several problems of *organization-internal* management of information security.

As it will be explicated in the following sections, the problem of information security *inside* organizations can be viewed from at least two further economic perspectives. First, organizational information security itself can be seen as an economic good that has some specific properties. These properties have to be taken into consideration to allow for a well-founded discussion of how information security is and how it could be realized inside organizations. Section 4.1 thus focuses these specific properties.

The second additional economic perspective on information security arises from the

[5] ISACs are associations that are set up to share security-relevant information among different (and in many cases competing) organizations. According to Gordon, Loeb, and Lucyshyn (2002), they can be compared to trade associations.

[6] A comprehensive overview of the research area has been given by Anderson and Moore (2006). For a more extensive overview, see Anderson and Moore (2007). See also the different papers presented at the yearly Workshops on the Economics of Information Security being available via `http://weis09.infosecon.net/past_workshops.html` [16.03.2009], for example.

[7] The reader might recall the principle of information asymmetries presented in section 3.2.2.

problem of making a cost-benefit analysis for information security already mentioned in section 1.5. The different types of costs and benefits arising from a specific "level" of information security are thus considered in more detail in section 4.2 to allow for a better understanding of what determines the payoff of information security efforts. Together, those two additional perspectives round off the first part of this work and lead over to the second one where the preceding general remarks on information security and the nature of organizations are merged into a single integrative framework.

4.1 Economic Properties of Information Security

This section gives a brief overview of those economic properties of information security that are most important within organizations. Even if most of these properties were originally mentioned for an inter-organizational context, they can also be mapped to the economic relationships between different players *within* an organization.

4.1.1 Information Security Externalities

The existence of externalities is one of the elementary principles of modern economics. In general, an externality can be defined as an *"uncompensated impact of one person's actions to the well-being of a bystander"* (Mankiw and Taylor 2006, p. 11). A quite common example for a negative externality is a plant polluting the air around it. People living in the same area do not have any influence on the decision to pollute the air but are at the same time affected negatively by the pollution. Thus, the fact of polluted air is a negative externality posed to the people living around the plant.

Positive externalities in turn exist where people profit from an action which they are not involved in. For example, if a person A gets vaccinated against a certain disease, not only A herself benefits from this vaccination but other persons do so, too: If A gets vaccinated, A will not get infected by the disease vaccinated against. This also lowers the chance for other people to get infected and thus provides a benefit to them. By this way, other persons profit from A's vaccination without bearing any costs at all.

It is now widely accepted that several aspects of information security are subject to negative as well as to positive externalities, too.[8] In most cases, such security externalities are considered in conjunction with inter-organizational dependencies[9] but such information security externalities are also present *inside* organizations. Consider, for instance, several people being connected through an organization-internal computer network. In this case, it could – under certain circumstances – pose additional risks to all others if only one of them behaves insecure.[10] By leaving his own computer

[8] See, for example, Schneier (2007b).

[9] See, for example, Rowe and Gallaher (2006), Kunreuther and Heal (2003) or Png, Tang, and Wang (2006).

[10] See, for example, Kumar, Telang, and Mukhopadhyay (2006, p. 2): *"[D]epending on the network architecture, it could take just one insecure network entity to put the entire network at risk"*. See also Kumar, Telang, and Mukhopadhyay (2007). Obviously, the presence of negative externalities is thus an economic perception of the well-known principle of the "weakest link". Note that this

insecure, person A thus causes costs – in the form of additional risks – for all other members of the network. To consider *Information* Security instead of simply *Computer* or *Information Technology (IT)* Security, think of different persons being bound together to perform activities that require using a common set of secret information or knowledge. If one person discloses this information, completely or partially, this would also cause harm – and thus, produce costs – to the others.

Positive externalities, in turn, do also exist in the world of information security. The easiest example, analogical to the vaccination example mentioned above, is a person A buying and installing a security software. Again, not only does A benefit from this decision but others do so, too: Their chance of getting a virus sent by A via email decreases as well as the probability of being affected by a (D)DoS-attack originating from A's computer. In this case, others profit from A using security software, whilst A bears the full costs.[11]

Externalities thus exist with regard to information security, too. They do not only exist between *different* organizations but also appear *inside* them and have thus to be taken into account when modeling organization-internal information security in an economic manner. We will return to this principle in more detail later in section 5.1, but as a first step, it is sufficient to recognize the general existence of externalities with regard to information security.

Conclusion:
Information security can be subject to positive as well as to negative externalities. Such externalities arise when the information security of one player depends on activities of others. This can also be the case *inside* organizations.

4.1.2 Information Security as Organization-Internal Public Good

Closely coupled with the economic term of positive externalities is the concept of "public goods". In economic theory, a public good is characterized by the two main characteristics of non-rivalry in consumption or use and non-excludability[12], where non-rivalry refers to the value and availability of the good not being decreased by its usage and non-excludability refers to no individual party being barrable from consuming or using the good.

These properties can also be assigned to the "economic good" of information security inside organizations.[13] If information security inside an organization is understood as

principle is not necessarily present in *any* network structure.

[11] These costs do not need to be limited to the costs of buying and installing the security software. Also, keeping the security software active could lead to lower productivity – and thus, costs – due to less connectivity or even slower systems. See also Kumar et al. (2006, p. 4): *"Each countermeasure [...] has costs in monetary terms, whether it involves purchasing products or results from productivity losses [...]."*

[12] See, for example, Mankiw and Taylor (2006, p. 208).

[13] The idea of information security as "organization-internal public good" was already examined by Glaser and Pallas (2007, pp. 3f). The considerations made herein are based on this article and expand the respective basic thoughts to a certain extent. See also Böhme (2005, p. 4) noting that

an economic good, then all members of the organization can profit from making use of this good – for example by being protected from external network attacks – without decreasing its value for others. Information security is thus non-rival in consumption or use.

Regarding the second characteristic of non-excludability, it can be argued that members of an organization could – in principle – very well be excluded from making use of the "economic good" of information security (by explicitly revoking protection from external network attacks, for example) and that non-members are of course excluded from making use of it. In this case, we had to consider "club goods" instead of "public goods". On the other hand, it can also be argued that an organization would not consider taking away information security from any of its members because of negative externalities possibly arising (see section 4.1.1 above) and that excluded non-members are not even part of consideration when thinking about information security *inside* organizations.[14] Furthermore, national defense is one of the classic examples for a "public good"[15] and the same restriction of a generally existing excludability applies here, too. For the time being, we will thus consider information security as an organization-internal public good.[16]

For public goods in general, the most important problem regards their production. As mentioned above, no one can be excluded from making use of public goods. The production of a good of this type thereby poses positive externalities to non-producers who are also able to profit from the good's production, leading to the producer not gaining all benefits arising from the good being produced. Resulting from this non-excludability and the positive externalities, no one would be willing to pay for the production of this good even if the individual benefit from the good being produced would outreach the individual portion of production costs having to be paid. Instead, every individual has an incentive to free-ride and use the good without paying for it. Consequently, this mechanism leads to no one paying for the production and to the good not being produced at all.[17]

To ensure the good's production anyhow, some instance or agreement is needed that forces all possible consumers of the good to contribute to its production. Here, classical economics suggest the production to be left to the government which then charges all

"network security appears to have properties of a public good".

[14] In fact, for the general classification it is only relevant if exclusion is *generally* possible and not if exclusion is likely to be realized. From this point of view, we had to consider "club goods" here. Nonetheless, Ostrom (2005, pp. 23 ff) clarifies that both attributes, rivalry as well as excludability, have to be considered as relative values ranging from low to high and not as dichotomized properties. Furthermore, Ostrom characterizes excludability by the *difficulty* of realizing exclusion and not by the mere possibility of doing so.

[15] See, for example, Mankiw and Taylor (2006, p. 210).

[16] As we will see later in chapter 10.1.4.3 there is also an alternative economic interpretation that presumably provides significantly more clearness and goes without having to decide between "public" or "club good". Nonetheless, current discussions usually consider – if economic aspects are taken into account at all – information security as a "good" that is demanded. We will thus proceed with characterizing this "good" in brief.

[17] See again Ostrom (2005, p. 24): *"When it is costly to exclude individuals from enjoying benefits from an investment, [private entrepreneurs] have few incentives to provide such services on their own initiative."*

citizens through taxes, for example. By doing so, the government can overcome the effect of free-riding and provide an overall benefit that outreaches overall costs.[18] In this case, any individual is forced to pay for the production of the public good, even if the costs might exceed the individual benefit in particular cases. None of the citizens has the choice of refusing to take part in production of the good.

The analogy from classic public goods to information security can be drawn by interpreting the considered organization as the equivalent of the state. Information security is the public good to be produced and any member of the organization has an interest in this good being present. Nonetheless, any single member has an interest not to invest in the production as he will – due to non-excludability – be able to profit from the good without paying for its production. This would ultimately lead to an outcome with no member investing in security at all.[19] Like governmental production of the public good is – combined with an enforced obligation for all individuals to invest in this production (taxation) – the usual economic way to overcome this dilemma, a possible approach for producing the good of information security would lie in the "government" of the organization producing the good and forcing all of its members to contribute to the "production costs". As we will see later in chapter 5.2, this is in fact the typical way for realizing information security *inside* organizations nowadays. Nonetheless, one can also think of alternative mechanisms for ensuring the appropriate "production" of the public good. These will be discussed in more detail in chapter 10.1.4.3. For the time being, it is sufficient to recognize significant similarities between information security inside organizations and the classical economic concept of public goods.

Conclusion:
Information security inside an organization features certain characteristics of a public good: Profiting from it is non-rivalrous as one member who is protected does not automatically lower the protection of another member. Furthermore, information security is non-excludable, meaning that no member can be barred from making use of it. These specific properties lead to information security not being produced by the members of the organization. Instead, the "government" of the organization might have to ensure the "production" of information security and enforce all members to contribute to this production.

[18] See, for example, Mankiw and Taylor (2006, pp. 209 ff) or any other economic textbook. Interestingly, Ostrom (2005, p. 24) uses the term *"toll goods"* for goods of low excludability and low rivalry.

[19] See also Kunreuther and Heal (2003), who analyze different comparable problem-settings of "interdependent security" (airline security, fire protection) and conclude that *"it is often advantageous for all agents to adopt protection for both themselves and society, but none of them have an economic incentive to do so on their own"*. Note also their explicit reference to the prisoner's dilemma.

4.1.3 Non-Measurability of Information Security's Value

Most problems with respect to the provision of public goods arise from the fact that it is often unclear what amount of the good should be provided to achieve an efficient outcome. If the government has to provide the public good of streets, for example, it has to decide where streets should be built, where not and of what kind the streets should be. It is in this case often unknown by the government where the value of a street would exceed the costs and where the possible benefits from a street would be too low for its construction to pay off later. Even if there are always cases where the benefits will clearly and unquestionably outreach the costs, there will at the same time be a multitude of cases with a significantly less clear cost-benefit relation. For these cases, it is usually impossible for the government to exactly determine the overall benefits that would be provided by the street and ultimately, the decision has to a large extent to rest upon *"rough approximations"* (Mankiw and Taylor 2006, p. 213).

A comparable but slightly different problem can be identified in the area of information security. If security is abstractly understood as the absence of risks[20], the value of information security could be "measured" indirectly by quantifying the risks being present.[21] But although there exist different and sometimes highly sophisticated methods for identifying and assessing information security risks, none of them allows for a precise and objective calculation.

The attack-trees suggested by Schneier (2000, pp. 318 ff), for example, are a powerful tool for identifying different possible attacks against a certain goal and for comparing the relevance of these different attacks. Attack trees thereby allow to determine, for example, the cheapest possible attack against a certain goal or the attack with the highest probability to occur. On the basis of such considerations, it is possible to calculate the value of the current state of information security and the potential value of a state that could be reached with certain additional countermeasures being in place. Attack trees could thus allow to estimate the value of different states of information security, thereby effectively "measuring" the value of information security. But even if attack trees allow for a more structured way of identifying and evaluating risks, they still depend on rough approximations and furthermore do not prevent from overlooking a cheap and viable attack, for example. With growing experience, these aspects will typically get less significant, but basically, attack trees still rest upon on a certain amount of intuition. As Schneier (2000, p. 332) himself puts it, *"there's always the chance that you missed an attack, but you'll get better with time. [...] [C]reating attack trees requires a certain mindset and takes practice."* Attack trees can thus be seen as powerful tool for structuring considerations, but they still require rather intuitive decisions that strongly depend on experience. They possibly lead to better estimations but they do not allow to really "measure" the level of information security or to determine the value of a specific level *objectively*.[22]

[20] See, for example, Blakley, McDermott, and Geer (2001)

[21] There is also the approach of "measuring" information security by means of "security metrics" (See, for example, Jaquith 2007). Nonetheless, such metrics only hardly allow for a specification of the *value* of a certain level or state of information security. Furthermore, the problems discussed below generally apply to security metrics, too (See Bellovin 2006).

[22] See also Schneier himself (2006, p. 5): *"The problem is that security's effectiveness can be extremely*

This is also true for different national and international standards concerning information security. The ISO standard 17799, for example, emphasizes the importance of risk assessment and contains a separate chapter for this issue[23] but nonetheless merely calls for a *"systematic approach"* to be used instead of providing detailed guidance on the risk assessment process.[24] The "Risk Management Guide for Information Technology Systems" published by the NIST (2002) only distinguishes between high, medium and low levels for the likelihood of an adverse event to occur as well as for the potential impact of such an event and gives no further advice about how to determine the levels of a specific threat. And finally, the German standard BSI 100-3 also does not even try to measure a level of information security or to determine a specific risk being present but rather gives advice how to structure the *process* of treating specific risks. The assessment itself and thus the determination of the current level of security is again basically left to intuitive or at least non-formal decisions.[25]

Overall, we can observe that established standards and practices do not assess the value of information security on a quantitative but rather on a qualitative basis.[26] Even if there exist different methodological frameworks for assessing information security risks and thus the value of a certain level or state of information security[27], the result of such assessments usually depends on intuitive estimations to a great extent. There is, as Anderson and Moore (2006, p. 610) put it, still a *"difficulty in measuring information security risks"* in a precise, comprehensive and accurate manner.

For the cost-benefit analysis having to be made with regard to the production of the "public good" of information security, this results in an inability to accurately define the "benefits" part of the analysis: As the benefit of information security is primarily the reduction of present risks,[28] the value of a certain level of information security is consequently as indeterminable as the risks themselves are. This does not necessarily have to lead to inadequate outcomes but it has to be accepted as economic property of the "good" of information security that its exact value can – at least with the methods currently available – in most cases not be determined exactly.

---

hard to measure."

[23] See ISO / IEC (2005a, pp. 5 f).

[24] In fact, another ISO standard (ISO / IEC 1998, ISO 13353-3,) is referenced as *discussing* different methodologies of risk assessment that could be applied. The different methodologies described therein do also not provide an objective measurement.

[25] In fact, the standard even negates the approach of explicitly considering occurrence probabilities because there are no reliable methods apart from mere speculation to assess them. See BSI (2005c, p. 3): *"[I]t has been proven that assessing the probability is often difficult in practice because there is no basis for reliable estimates. The interpretation of the probability is also frequently questionable."*

[26] This was also observed by Blakley et al. (2001, p. 99): *"The large majority of [standards regarding the analysis of information security risk] have been qualitative – that is, their assessment of probability and consequence of risks is based on a 'low/medium/high' characterization rather than on a specific probability and a specific dollar amount of loss."*

[27] Besides those mentioned here, many other approaches exist for estimating information security risks. For more extensive considerations, see for example Peltier (2005, esp. pp. 41 ff) or Alberts and Dorofee (2002, pp. 169 ff).

[28] There are also other benefits which are outlined in section 4.2 in more detail. But basically, these can also not be determined in a precise manner and are thus subject to comparably vague estimations, too.

Conclusion:
Like it is typically the case for public goods, the value of the "good" of information security can hardly be measured or determined in a quantitative manner. Established standards and practices usually assess information security's value in a *qualitative* and not in a *quantitative manner*. Furthermore, these assessments have in most cases to be made on the basis of intuitive considerations to a certain extent, thereby only allowing for *rough approximations* of the value being provided by a certain level or state of information security.

4.2 Information Security Payoff for Organizations

If information security can – as outlined in section 4.1.2 – be seen as an organization-internal public good that has to be produced by the "government" of the organization, the question is which amount of this public good should be produced to reach an efficient outcome. Even if the value of information security can, as outlined above, not be measured accurately, one can nonetheless make theoretical considerations about this value or payoff of different amounts of security being in place.

Usually, it is argued that a higher degree of security lessens the faced risk and thereby increases an organization's overall profits up to a point of the optimal level of security. From this point onwards, more security still lessens risk but also has negative effects (the reduction of work efficiency, for example) that over-outweigh the positive effect of risk-reduction, leading to a counterproductive effect of "too much security". Using economic terms, the optimal level of security is reached where the marginal costs of an additional unit of security equal its marginal benefit. Björck (2001, p. 1) puts this effect into words as follows:

> *Too much business security will increase [...] costs and reduce [...] potential revenue streams substantially [...]. The goal of security management in organisations should therefore be to identify and strive toward the optimal point between security and insecurity.*

These effects of changing the level of information security are represented in figure 4.1. Even if this *"security payoff curve"*[29] – as it will be named herein – is not explicitly examined in more detail by Björck, some factors determining the payoff for a certain level of information security and thus defining the shape of the payoff curve can easily be identified.[30]

[29] Some readers might expect the term of a ROSI (Return on Security Investments) to be used here. However, this term is potentially misleading in a certain manner and will thus not be used herein. See, for example, Bejtlich (2007), taking up the position that security does not create wealth but only avoids the loss of wealth and can thus not have any *return*: *"The key principle to understand is that wealth preservation (saving) is not the same as wealth creation (return)."* See also Schneier (2008) as well as Shostack and Stewart (2008, p. 119).

[30] The considerations should be understood as being idealized and not as trying to perfectly represent reality. The unidimensional and continuous nature of curves like those used herein might be misleading in a multitude of ways. However, it is the aim of these considerations to identify

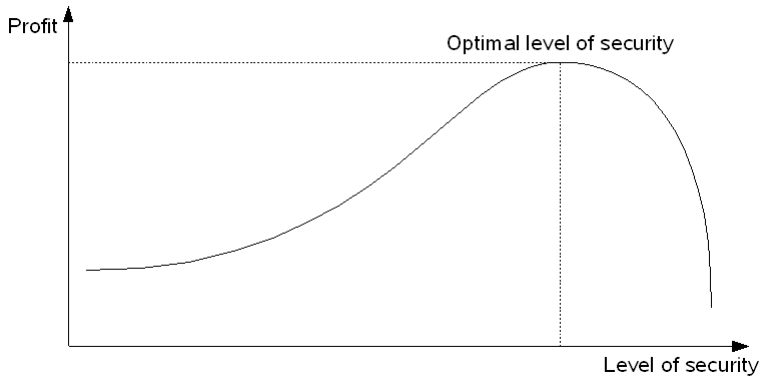


Figure 4.1: Optimal level of information security (according to Björck 2001, p. 1)

First, there are *direct costs* of Information security. These refer to the costs of technical security solutions as well as to the salaries for security personnel operating these solutions, the costs of conducting awareness trainings or the expenditures for establishing and enforcing a system of formal information security rules. Slightly more formally than Björck (2001), we can outline these direct costs as follows:[31]

The level of security is assumed to be a value between 0.0 and 1.0 with 0.0 representing a state of "no security at all" and 1.0 standing for the (imaginary) state of "absolute security". Within this range, direct costs of information security are zero for the level of no security, of course. Starting from there, heightening security can initially be realized at relatively moderate costs.[32] But with higher levels of security already being in place, it is a truism that every "additional unit" of security involves significantly increasing costs. With the assumption of the organization always trying to get the "next unit of information security" for the lowest possible cost, it becomes evident that the marginal direct costs having to be paid for each additional "unit" of information security are constantly increasing, resulting in a constantly increasing gradient of the "direct costs" curve. Of course, the level of absolute security can never be reached and the direct costs for approaching a level of *nearly* absolute security would be gigantic. As a rule of thumb, it can be expected that every halving of the remaining risk leads to a doubling of direct costs.[33]

As a second factor of the security payoff curve, we can identify *indirect costs* of information security. These costs refer to all adverse effects of information security,

general principles and for this purpose, such curves are particularly suitable. Furthermore, it should be noted explicitly that Björck's model – as well as most other models – basically assumes a state of total insecurity as default and thus as initial point for all further considerations.

[31] I am grateful to Kai Dietrich for fruitful discussions about the actual shapes of the different sub-curves.

[32] Even the installation of a simple firewall or a virus scanner can, for example, eliminate a wide range of threats that would otherwise pose considerable risk.

[33] See, for example, Adi Shamir in his Turing Lecture of 2002: *"To halve your vulnerability, you have to double your expenditure"* (slide 8).

most of which arise from impairment of daily work practices resulting in a lower ability of the organization to generate value. Due to its nature of *"prevent[ing] people from doing something"* (Schneier 2006, p. 13), security can in certain cases also prevent people from doing something expedient that would generate value for the organization or from doing this efficiently.

Even if these indirect costs cannot be formalized as intuitively as the direct costs, some educated guesses can be made: Like direct costs, indirect costs resulting from hindered value generation are non-existent for no security being in place. And like with direct costs, the first steps toward a higher level of security will typically result in rather insignificant impairments of daily work and thus lead to nearly negligible indirect costs. But with a certain level of information security already being present, additional measures will start to disturb efficient work practices, necessitate additional procedural steps having to be executed or even bother the respective users. At this stage, heightening the level of information security begins to prevent substantial amounts of value generation and the respective subcurve thus has to increase materially. Nonetheless, these indirect costs can never reach or even exceed the amount of value that could be realized in general. No security measure can prevent more value generation than would originally be possible without security being in place. With the level of security increasing further, the respective subcurve will thus approach the level of value generation that would be possible with no security being in place.[34]

Besides direct and indirect costs of information security, there are also benefits, of course. And again, these benefits can be direct or indirect.

Direct benefits are herein understood as the reduction of risks faced by the organization. If a risk can be eliminated by a certain security improvement, it is this risk reduction that represents the direct benefits from information security. Risk, in turn, is usually understood as the mathematical product of probability of a loss occurring and the potential loss. As the probability of occurrence can be assumed as being inversely proportional to the level of security and with potential loss being constant, direct benefits of information security rise linearly with an increasing level of security. With no security being in place at all, direct benefits are zero and with "absolute security", direct benefits would due to a non-existing probability of occurrence reach the maximal potential loss of the organization.

And finally, the payoff of information security is also determined by *indirect benefits*. These benefits have to be considered especially in conjunction with certifications of a certain level of security being present inside an organization. As mentioned earlier (see section 2.2.4), potential contractors can, for example, require compliance to a certain information security standard for starting business connections at all or for the permission to participate in an invitation to tender. In this case, a certain level of information security opens up an area of business opportunities for the organization that would otherwise not exist. Better credit conditions because of Basel II regulations have also to be seen as such indirect benefits.

[34] Additionally, it could be argued that the possible indirect benefits that will be outlined below had to be added to this maximum. Nonetheless, the indirect costs would in this be limited to a certain maximum, too.

Like all other subcurves, these indirect benefits of information security will begin with a value of zero for no security being present and rise very slowly with an increasing level of security. But at a certain critical point the required level for established security certifications (on the basis of the ISO 27000 series, for example) will be reached and this level has characteristics of a critical mass: Once reached, the resulting benefits increase abruptly when the first certification can be realized and market opportunities are broadened significantly. On the other hand, when one certification is achieved, the additional value of further certifications will be less significant. From a certain point onwards, additional indirect benefits can thus hardly be derived from a further increase in the level of information security – as all relevant certifications leading to additional business opportunities are already achieved, for example. The subcurve for the indirect benefits will thus – similar to the indirect costs curve – be S-shaped, even if the rise is more abrupt. The maximum of this subcurve is defined by the amount of new business opportunities resulting from certifications, for example.

These four elements define the overall security payoff curve for the organization, where direct and indirect costs are considered as negative factors, of course. Figure 4.2 gives an exemplary graphical representation of the four subcurves. When these curves are added together, they actually result in a graph that differs from the one being drawn by Björck (2001) only slightly.

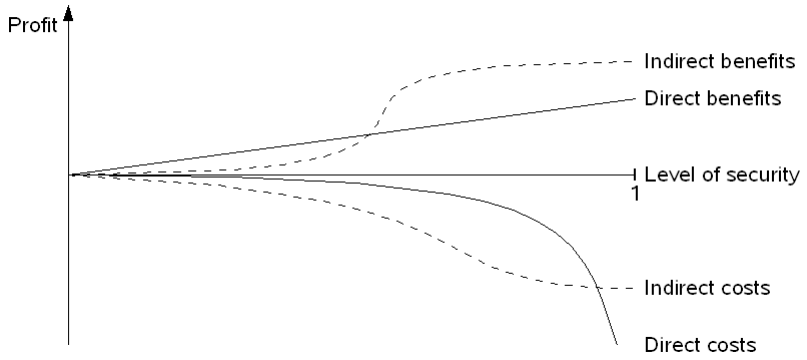


Figure 4.2: Relevant factors for security payoff curve – Generalized illustration

By identifying four subcurves with different specific characteristics, we thus specified the rather intuitive model of Björck (2001) in slightly more detail. We identified the overall payoff of information security for organizations to consist of four factors, namely direct costs, indirect costs, direct benefits and indirect benefits. If, as Björck (2001, p. 1) stated, *"[t]he goal of security management in organisations [is] to identify [...] the optimal point between security and insecurity"*, security management thus had – from an abstract point of view – to be aware of the specific payoff curve and thus of the four specific sub-curves for the considered organization to derive an optimized level of security that should be realized.

Nonetheless, the model of the four subcurves is still largely abstract and the actual payoff curve for a certain organization will typically be as individual as the organization

itself. For example, the importance of the indirect benefits might be nearly non-existent for a certain organization while another one might not be able to exist at all without the respective certifications.[35] For some organizations, the indirect cost curve might begin to rise significantly at a comparably low level of security while others will hardly suffer any productivity losses before a substantive level of security is reached. And finally, even within one and the same organization, different divisions and individual members will – because of different functions and job profiles – typically face different payoff curves, too.

Due to this multitude of factors influencing the payoff of a given level of information security, determining some kind of an "optimal" level or state of information security gets even more complicated. It is not only the direct benefits that can, as outlined in section 4.1.3, not be assessed as precise quantitative values. There are also various kinds of indirect benefits and indirect costs which will in most cases be equally non-measurable and non-predictable. Similar to the direct benefits, the valuation of these factors will thus have to be strongly based on intuitive and rough approximations, too. As we will see later in chapters 6 and 9, this effect strongly gains importance with ongoing changes of organizational structures.

Let us herewith close our examination of established economic perspectives on information security. So far, we have discussed some economic properties of information security and identified four distinct components defining the overall payoff of a certain security level for an organization. Even if these discussions already referred to economic concepts and principles, however, they have so far not yet been linked to the economic foundations of organizations as outlined in chapter 3. The same is true for the non-economic considerations regarding the current state of information security inside organizations made in chapter 2: Even if the different basic principles covered in this part are – from different perspectives – all relevant for *information security inside organizations*, we have so far not even started to interlink them with each other in order to gain new insights. The following part shall therefore establish these still missing links in order to derive a first stage of an integrative, economically inspired, positive model of information security inside organizations as outlined in section 1.5.2.

Conclusion:
The payoff of a certain level or state of information security is determined not only by the risk being present but rather consists of at least four factors: direct costs, indirect costs, direct benefits and indirect benefits. The idealized subcurves for these factors show some distinguishing characteristics leading to an overall payoff curve for information security that slowly rises upwards to an "optimal" level of information security and from there on falls steeply. The multitude of factors determining the overall payoff curve for information security leads to decisions that have strongly to be based upon intuitive estimations and approximations instead of relying on precise calculations.

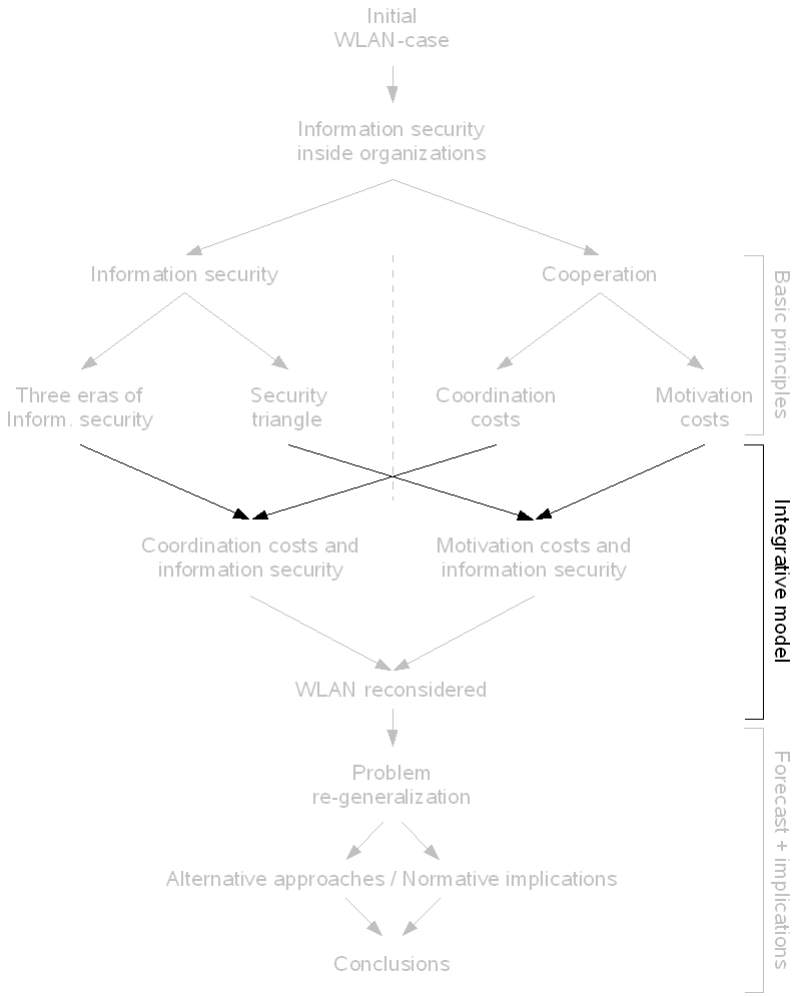
[35] Consider, for example, an organization selling items online. For being able to accept payments via credit cards, the organization has to comply with certain standards. For this specific case, the indirect benefits are much more important than for an engineering organization, for example. See also Shostack and Stewart (2008, pp. 106 ff).

Part II

Information Security and Costs of Cooperation: Toward an Integrative Model

Chapter 5

Information Security, Cooperation and the Hierarchical Approach



Chapter 5

Information Security, Cooperation and the Hierarchical Approach

> *The only way to erect such a Common Power [...] is, to conferre all their power and strength upon one Man, or upon one Assembly of men, that may reduce all their Wills [...] unto one Will;*
>
> – Thomas Hobbes

The beginning of a new part gives us the opportunity to pause for a moment, look back and reflect what we have accomplished so far. After depicting the concrete problem of publicly usable WLAN in chapter 1, we reviewed the current state of information security from two different abstract perspectives and developed the two distinct models of *"three eras of information security"* (section 2.1) and of the *"security triangle"* (section 2.2).

In chapter 3, we elaborated the economic basics of cooperation among different individuals, identified two general models for realizing this cooperation – markets and hierarchies – and worked out how the cost structures of these models are, how they can be affected by technological change, and what the role of hybrid models lying between the two idealized extremes is. In particular, we explained that any kind of cooperation has to solve the two tasks of coordination and motivation, both of which entail specific kinds of costs – *coordination costs* and *motivation costs*.

And finally, we went into the details of some selected aspects of information security that are strongly related to well-known economic principles (chapter 4). Leaving aside the latter security-economic aspects which will be of general relevance throughout the remainder of this work, we can depict the structure of this work so far as done in figure 5.1.

With these basic considerations, we have gathered all necessary foundations to develop a consolidated model of information security inside organizations that associates the fundamental economic principles of organizations with established methods of organization-internal information security. It is thus the aim of the following chapters to bring together the different aspects which have so far been considered separately. We shall thereby derive first ideas of an "organization-economic view on information

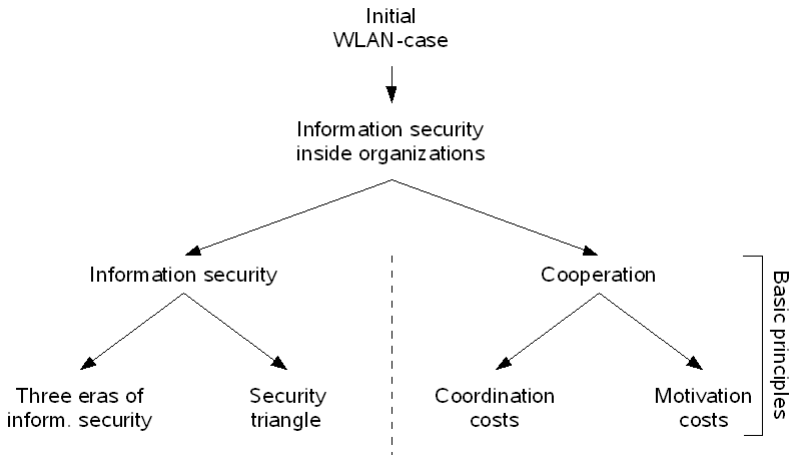


Figure 5.1: Structural overview of preceding aspects.

security". To do so, let us return to our introductory problem of WLAN-security in a first step and show that the underlying problem is basically one of *cooperation*.

5.1 Information Security as Cooperation Problem

As shown in chapter 1.2, publicly usable WLAN poses a considerable challenge for organizational information security. On the one hand, the use would allow an organization to generate value, but on the other hand, usage would also introduce additional risks. Four general approaches for tackling the underlying conflict of goals were identified of which one – the change of the underlying hotspot infrastructure – is currently unrealistic. The three remaining approaches – a general ban, login automation and lax treatment – were all shown to entail benefits as well as costs for the organization.

From an organization-internal view, all these approaches represent attempts to influence the behavior of the different individual members of the organization. No single player can decide on his own to use WLAN or not but is rather constrained in his decision by preventive technical means, by a reduction of hotspots being usable or by organization-internal rules obligating him to establish a VPN as soon as possible, for example. In the outcome, all these constraints lead to different results than uninfluenced decisions of the individual members would.

The reason for such constraints of individual behavior to exist is the principle of negative externalities delineated in section 4.1.1: Basically and without any of the established countermeasures in place, the behavior of any single member with regard to the use of public WLAN would strongly affect all other members and thus the organization as a whole. Assume, as a very simple example, an organization consisting of ten members with one of these ten – member A – having to decide if he uses a public WLAN hotspot or not. Assume furthermore that using the hotspot would allow A do derive a profit of 10 for the organization (by closing a contract, for example) and

that WLAN usage would at the same time result in a calculated risk of 5 arising from the possibility of A's notebook being attacked. For this isolated case, it would be economically rational to use the hotspot as the resulting profit exceeds the resulting risk.

But due to the existence of externalities, A's decision to use the hotspot would also affect the remaining nine members of the organization. For the case of A being successfully attacked, all of them had to bear the risk of being attacked over the internal network of the organization, of some kind of collective internal resource being compromised or of the confidentiality of some organization-internal information being breached, for instance. Assume that this results in a calculated risk of 1 for each of the remaining nine members. In this case, the organization as a whole had to bear an overall risk of $5 + (9 \cdot 1) = 14$ while the overall benefit of A using the hotspot would still be 10. From a collective point of view, connecting to the hotspot would thus result in a negative payoff of -4 and consequently, the organization as a whole had an interest in A not doing so. Even in this simple example, the balancing of risks and benefits would thus lead to inefficient outcomes for the organization as a whole if it were realized by the different individual members alone and without any kind of collective considerations being made.[1]

Individual interests make things even worse. Assume that A could not only derive a benefit for the organization from using the hotspot but were also able to derive an *individual* profit from doing so. This individual profit could result from time savings as well as from an individual bonus for closing the contract, for example. Newer management approaches like management by objectives (MBO) might also provide individual incentives.[2] In this case, A would possibly have an interest in using the hotspot even if he knew about the negative externalities being imposed on the other members and thus on the organization as a whole. Furthermore, all other members would possibly make their individual decisions in the same manner resulting in a situation where every single member is worse off than he would have been if all members had abstained from using hotspots.[3]

The same scheme of individual decisions possibly leading to undesired results for any member of an organization is nearly omnipresent in the field of information security. Using an external storage device to take organization-internal data home can pose a

[1] Of course, there are also cases where using the hotspot would very well be in the collective interest. For an "organization" of two members, for example, the same values of 10, 5 and 1 would lead to a positive overall payoff of $10 - 5 - 1 = 4$ from the hotspot being used by one of the members. If the realizable benefit would represent a value of 20 instead of 10 in the ten-member-organization, the benefit of A using the hotspot would also exceed the calculated risk and using the hotspot would again be in the collective interest of the organization.

[2] For an introduction to management by objectives, see Drucker (1954, pp. 121 ff). Generally speaking, management by objectives primarily sets goals to be reached by the person being "managed" and gives no or only little advice on how to reach these goals. Furthermore, payment is in many cases also based on the achievement of the objectives and not on variables like time spent for working alone. Obviously, this entails additional incentives for the individual to behave in a way that minimizes individual effort for achieving the objectives and to be less concerned about possible negative externalities imposed on others.

[3] For vivid examples of the outcomes of such problems of collective action in the field of security, see Kunreuther and Heal (2003).

benefit to a single member, but if this device is lost, internal data is possibly made public and can result in substantial losses for the organization as a whole and thereby lead to considerable loss for any of the members. Even if it is known by every member that the overall effect of taking internal data home on external devices is unquestionably negative, any single member has an individual interest in doing so. Without any mechanism fostering cooperative behavior being in place, this would again result in a situation with all members taking internal data home and with all members consequently being worse off than they would have been if they had a mechanism that realized cooperation.[4]

And finally, even positive externalities can also have the effect of leading to suboptimal outcomes without cooperation. If user A from the exemplary ten-member-organization above bought and installed an antivirus-software, this would result in a certain individual benefit of, say, 5 but at the same time entail costs of 10. Buying and installing the software would thus be irrational for A because the costs would exceed the benefits. But if all the remaining members also profit from A installing the software with a benefit of 1, the overall benefit would be 14 while the overall costs would still be 10. From a collective perspective, it would thus be preferred that A bought and installed the software. Again, individual and collective interest diverge because of externalities.[5] And again, the advantageous overall outcome can only be realized through cooperation.[6]

Of course, these examples are somewhat simplified, but they nonetheless illustrate the general need for cooperation among different members of an organization in matters of information security to reach efficient outcomes for the organization as a whole. In most situations, the behavior required for achieving efficient collective outcomes differs from the behavior that would result from individual decisions and consequently, a mechanism is needed to realize the necessary cooperation.[7] As we will see later in section 7.2, the different mechanisms for constraining the use of WLAN hotspots or external storage devices serve exactly this purpose: They are at least part of a wider strategy for realizing cooperative behavior among the different members of the organization to avoid unwanted outcomes.

---

[4] Some readers might see a resemblance to Hardin's (1968) *"Tragedy of the Commons"*. This similarity is not accidental but rather noteworthy.

[5] As the similarity of values suggests, the case of positive externalities does not differ much from that of negative externalities above. The positive externalities from buying the software could also be interpreted as negative externalities arising from not doing so. The only difference is the state being considered as the initial one.

[6] Some readers familiar with the concept of "internalizing externalities" (see Coase 1960, for example) might be tempted to argue that there could be some kind of compensation mechanism between the acting member and those affected by the externalities. This would, however, also represent a mechanism for fostering cooperation. Here, it is considered how the situation would be without *any* such mechanism being present. The Coasian concept will, however, be taken up later in section 10.2.1.

[7] Note that Bowles (2004, p. 25) explicitly mentions externatities as major reason for the necessity of mechanisms that foster cooperation: *"The reason why uncoordinated activities of individuals pursuing their own ends often produce outcomes that all would seek to avoid is that each person's actions affect the well-being of others [...]."* Note also that Bowles uses the term of "coordination" for what is termed "cooperation" herein.

One of the most important goals of realizing information security inside organizations is thus to establish cooperation among the different members in order to reach preferable overall outcomes. Even if this insight might seem trivial at first sight, it is essential for the subsequent considerations which are based on the theoretical examinations from the previous chapters: We concluded above that cooperation can – as a matter of principle – be realized through *different* mechanisms ranging from strict hierarchies over different kinds of hybrids to pure markets all of which have different cost structures. Of these different options, current practices of information security inside organizations are, as section 5.2 will show, mainly of hierarchical nature. The subsequent chapters will thus examine the "costs of organizedness" arising from hierarchical cooperation in the field of information security.

In doing so, coordination costs and motivation costs will, congruent with the general model of hierarchical cooperation developed in section 3.2, be considered separately. As we will see, coordination costs are strongly influenced by the respective computing paradigm being used. We will thus examine the security-related costs of coordination in relation to the model of the three eras in chapter 6. As we will also see, the different "meta-measures" from the model of the "security triangle" are basically aimed at influencing behavior in a specific manner. They thus represent mechanisms of motivation and consequently, security-related motivation costs will be analyzed with regard to the "security triangle" in chapter 7.

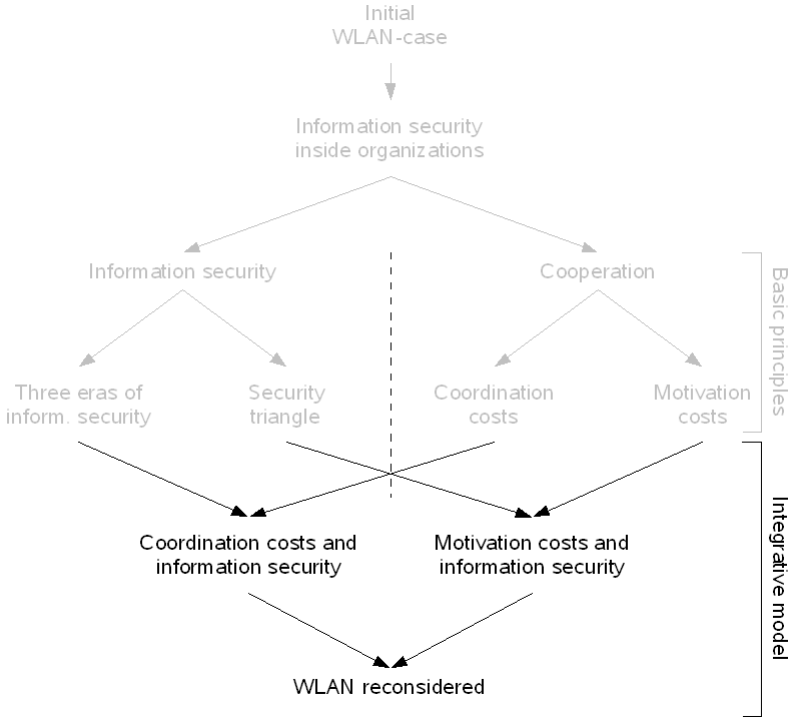


Figure 5.2: Structural overview of preceding and subsequent aspects.

With coordination and motivation being considered this way, we will then have developed a first version of an economically inspired, positive model of organization-internal information security. Based on this model, chapter 8 will reconsider our initial problem of public WLAN hotspots from section 1.2. Figure 5.2 visualizes the underlying logical structure. Part III will then discuss how current changes in information technology and organizational practices affect the cost structures being present in the field of information security inside organizations more generally and reflect about some possible approaches for lowering these foreseeable costs.

Conclusion:
Due to the existence of externalities in the field of information security, individual decisions alone would lead to outcomes that contradict the collective interest of the organization. Realizing information security thus necessitates cooperation among different individuals. This cooperation has to be fostered by appropriate mechanisms.

5.2 Information Security Management as Hierarchical Practice

As pointed out above, information security inside organizations depends on cooperation among different involved individuals. As possible approaches for realizing such cooperation, we identified the two idealized models of markets and hierarchies in chapter 3.[8] If, then, information security basically requires cooperation and if cooperation can basically be realized through markets and through hierarchies, it seems desirable to clarify in which way cooperation among different members of an organization is currently realized for the field of information security. Only then will we be able to examine the economic principles of organization-internal information security in more detail and to derive theoretically well-founded statements.

Current practices of information security inside organizations are generally – and implicitly in most cases – hierarchical in their very nature. Information security is typically *managed* by some *central* IT- or risk-management function which determines the course of action for the whole organization. This is for example done by formulating the overall information security strategy, by composing an information security policy, or by deriving the respective concrete security procedures in the form of technical countermeasures, formal security guidelines and more informal activities like conducting awareness campaigns. All these activities are typically planned and implemented by some central instance possessing absolute authority for the field of information security inside the respective organization.

The rather intuitive observation of information security currently being a hierarchical practice in most cases can also be drawn from a wide variety of existing literature. For example, the presumably most authoritative document in the field of

[8] Even if there is the possibility of realizing cooperation through different kinds of hybrids between those extreme cases, we will during our first steps only consider these idealized types for demonstrative purposes.

organizational information security – the ISO 17799 (ISO / IEC 2005a)[9] – calls for a *"management framework"* to be established inside organizations to *"initialize and control the implementation of information security"* (p. 9). The management is asked to *"provide clear direction [...] for security initiatives"* and should *"formulate, review, and approve"* an information security policy for the organization (ibid.). This policy document should be *"published and communicated to all employees"* (p. 7) and be subject to periodical *"management review"* (p. 8).[10] Even if it is left open by the standard if these functions should be adopted by a dedicated and specialized information security management or by the general management, these and other wordings implicitly allude to a hierarchical treatment of organizational information security.[11]

Closely coupled with ISO 17799 is the ISO standard 27001 (ISO / IEC 2005b), which is explicitly aimed at *"Information Security Management Systems (ISMS)"*. This standard requires an organization to establish an ISMS *"within the context of the organization's overall business activities"* (p. 3). Like in the ISO 17799, management is stipulated to establish a policy, provide ressources, decide on the acceptance of risks, and conduct regular reviews with regard to information security and the respective ISMS (ISO / IEC 2005b, pp. 9 ff). All these suggestions are also pointing toward a hierarchical treatment of information security inside organizations. The standard furthermore suggests that information security inside organizations is to be realized through management and that this management should follow a process approach based on the *"Plan-Do-Check-Act (PDCA) model"* (pp. V f). Even if not explicitly stated so, the PDCA-approach implicitly assumes that information security is realized in a continuous and well-defined process of conscious overall planning, implementation of the "plan", review of the achieved status quo, and refining. Especially the idea of conscious overall planning is an additional sign for the underlying assumption of centralized hierarchical practices.[12]

The German standard BSI 100-1 (BSI 2005a) also suggests information security to be realized in a hierarchical manner. It also refers to the PDCA-model being used in the ISO standard 27001[13] and understands information security management as

> *"the term used for the planning and supervisory functions that are required to assure the meaningful development, practical feasibility and effectiveness*

[9] The standard is explicitly aimed at establishing *"guidelines and general principles for initiating, implementing, maintaining, and improving information security management* in an organization" (ISO / IEC 2005a, p. 1, emph. added)

[10] Note that "management" shall not simplistically and generally be equated with a hierarchical approach. It is not the term "management" that leads us to the conclusion of current practices being mainly hierarchical but rather the context in which it is used within the mentioned documents.

[11] The same is true for the issue of risk-assessment and risk-treatment, which should, according to ISO 17799, follow a *"systematic approach"*. Furthermore, the scope of risk assessment can – besides others – be *"the* whole *organization [or] parts of the organization"* (ISO / IEC 2005a, p. 5, emph. added), also suggesting a centralized risk-assessment.

[12] Furthermore, the graphical representation of this process and the respective explanations (ISO / IEC 2005b, pp. V f) are at least noteworthy, suggesting that the PDCA-process would *"transform"* the *"input"* of *"security requirements"* into the *"output"* of *"managed information security"*.

[13] See BSI (2005a, pp. 14 f).

> *of a well thought-out and systematic IT security process as well as all the IT security measures required for this" (BSI 2005a, p. 16)*[14]

Again, conscious planning and a systematic process are central aspects of the approach. The *"topmost management level"* is considered to be *"responsible for assuring IT security"* and to have to *"actively initiate, manage and supervise the IT security process"* (ibid.). The complementary document on the methodology of the BSI's *"Grundschutz"*-approach (BSI 2005b) also suggests information security to be realized in a hierarchical manner. For example, it recommends different options for organizational structures to be used in different kinds of organizations, all of which mention an IT Security Officer who *"co-ordinates and drives IT security forward in the organisation"* (BSI 2005b, p. 19) and who is *"responsible for managing* all *the IT security issues within the institution"* (ibid., p. 20, emph. added).

Without going more into detail, we can thus – in spite of the typical highly abstract wordings used in these standards to be applicable to as many cases as possible – identify an underlying implicit assumption for ISO 17799 as well as for ISO 27001 and the BSI 100-x standards that information security should be *"managed"*[15] and that this should basically be done in a centralized, hierarchical manner. Even if not being stated so explicitly, all standards assume that there is one instance – be it a single "Chief Information Security Officer (CISO)" or an information security team – within the organization that is responsible for the overall information security and that this instance performs the different tasks of assessing information security risks, setting up an information security policy, planning countermeasures and propagating these downwards through the whole organization.[16]

Existing standards for information security inside organizations thus follow the classical management approach of conscious planning based on extensive information and of implementing the developed plans throughout the organization by hierarchical means to achieve cooperative behavior among the members.

Our known example of using public WLAN shall also illustrate the hierarchical nature of current information security practices. As outlined in section 5.1, the use of public WLAN by single members of an organization is subject to a problem of cooperation. If individual members would not cooperate and would follow their individual interest instead, the overall outcome for the organization would – due to too many risks being taken – be worse than it could be with cooperation in place. Organizations

[14] Note that the BSI uses the term of IT-security even if it is aware that "information security" would fit better. See (BSI 2005a, p. 8): *"The term 'information security' instead of IT security is therefore more comprehensive and more appropriate. Since [...] the term 'IT security' is still predominantly used in the literature (among other reasons, because it is shorter), it will continue to be used in this publication"*

[15] In fact, both ISO standards as well as the BSI 100-1 also carry the term *"management"* in their title.

[16] This conforms with the perception being propagated by a wide variety of business information as well as with usual experience in the field of organizational information security. Revealingly, there is no well-established standard textbook on the issue that would do more than ultimately reproducing the different best-practice-standards and guidelines. Nonetheless, those textbooks existing also expressly suggest a hierarchical top-down approach. See, for example, Whitman and Mattord (2003, pp. 20 f).

therefore establish mechanisms to realize cooperation and in fact, they typically do this in a hierarchical manner.

Following the approach outlined above, an organization will presumably have a single instance being responsible for information security throughout the organization – a security team, for example. Regarding the use of public WLAN, this security team will in a first step consider the pros and cons of this use. They will thus try to estimate the benefits that could be derived and assess the risks that would result for the organization as a whole. Assume that these calculations are consistent with those being made in section 5.1. The security team will then conclude that it would be advantageous for the organization as a whole if WLAN were not used. The non-use of WLAN would thus be the desired outcome of cooperation within the organization.

To achieve this outcome, the security team will then use different countermeasures. It might, for example, incorporate the use of WLAN into the overall security policy of the organization, ordering all members not to use public WLAN hotspots and threatening sanctions for the case of noncompliance. Obviously, this would be the prime example for a hierarchical relation between the security team and the "ordinary" member, as the security team would issue an order that had to be obeyed.

Technical countermeasures like the "general ban" supported by an "enforced VPN" mentioned in section 1.3.1 could also be instituted by the security team to achieve the desired outcome of members not using public WLAN hotspots. Again, this would be a hierarchical realization of cooperative behavior as the security team would enforce the according technical solutions to be used throughout the organization "from above". Members would have no choice but are rather coerced into conformance to the behavior assigned to them by the security team and enforced through the technical solution. Again, it is obvious that this constitutes a hierarchical relation between the security team and the "ordinary" member.

Finally, the security team could also conduct awareness trainings to emphasize the negative effect of using public WLAN. Members could be informed about the resulting overall risks or the infringement of collective interest and they could be appealed not to use WLAN hotspots. Such awareness trainings and comparable approaches are more sophisticated to classify than the methods mentioned so far. Typically, they are not directly aimed at implementing an exactly pre-defined or pre-planned behavior but rather try to achieve "secure behavior" on a more general basis. But basically, the usual applications of awareness trainings and comparable measures inside organizations can also be deemed as being hierarchical, as the aims and the contents of the training are primarily defined by the security team and as members should be influenced to behave in a manner that had been determined desirable in advance. Generally speaking, members ought to follow the intentions of the security team instead of acting in the way that would maximize their individual benefits. Even if the hierarchical nature of awareness trainings is less clear-cut than for the security policy and for technical approaches, such measures can thus also be considered as mechanisms of hierarchical influence that is exerted by the security team on the "ordinary" members of an organization.

Established mechanisms for realizing the necessary cooperation among different members with regard to information security inside organizations are thus hierar-

chical in their very nature. Additionally, different standards regarding organizational information security also suggest to pursue hierarchical approaches of managing information security to reach desirable outcomes. Hierarchy can thus be seen as the currently prevailing model for solving the cooperation problem of information security inside organizations.

If this is the case, then economic theory as outlined in section 3.2 teaches us that we have to expect this model of cooperation to entail *hierarchy-costs* in the form of

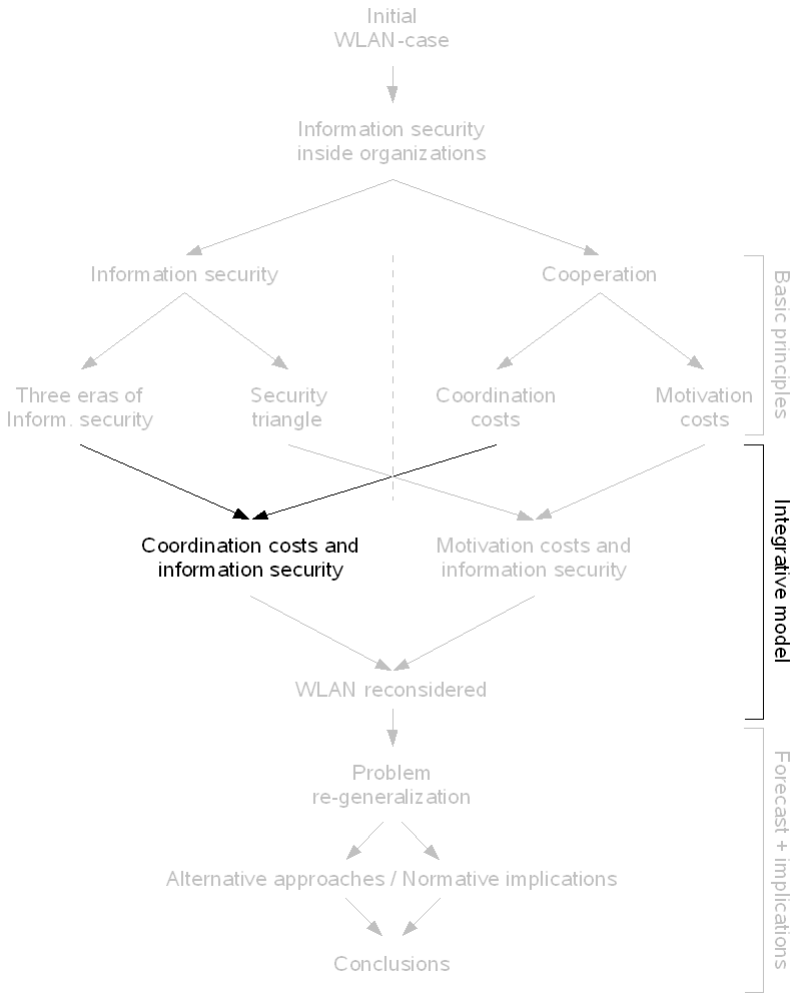
- *Coordination costs* for determining the allocation of resources and activities that is necessary for achieving the best possible outcome and
- *Motivation costs* for making the different players actually behave in the way that was determined during the process of coordination.

These different kinds of costs are often confused in the field of information security and they will thus be considered in more detail in the following. In doing so, we will retain the intuitive order of examining the task of coordination and the respective costs first (chapter 6) and the task of motivation and the adherent costs subsequently (chapter 7). For both kinds of costs, we will furthermore consider the influence of technological development whenever relevant. In particular, some interrelations between implications of technological progress for the task of coordination on the one hand and the change of motivational instruments being used on the other will be discussed in section 7.4. As it will become clear, technological progress had significant direct as well as indirect consequences for the cost structures of organization-internal information security.

Conclusion:
Established standards suggest that security-related cooperation inside organizations is basically realized in a hierarchical manner. This conforms with the usual practice: Typically, one instance is responsible for the information security of the whole organization. This instance performs the task of coordination and enforces cooperative behavior by hierarchical means. Hierarchy-costs can thus be expected to arise and should therefore be considered consciously.

Chapter 6

Information Security and Hierarchical Coordination Costs



Chapter 6

Information Security and Hierarchical Coordination Costs

In the case of a machine-shop which is managed under the modern system, detailed written instructions as to the best way of doing each piece of work are prepared in advance, by men in the planning department

– Frederick W. Taylor

As outlined in chapter 3, the task of coordination generally refers to the problem of identifying which activities should be conducted by which players and how other resources should be allocated to achieve an efficient outcome. Coordination thus tries to identify the state that would provide the best overall result.

For information security, this coordination especially refers to the access[1] to different organization-internal and -external resources by the different members and non-members or more abstractly, to the determination of the optimal state of security-related member behavior.[2] Notably, the outcome of this coordination process does not consist in statements like *"A technical solution has to be installed that prevents members from using public WLAN"* or *"File access rights have to be set in a manner that prevents non-members of a certain project from accessing files of that project."* Rather, the coordination process results in preliminary conclusions like *"Members should not use public WLAN (because this would be contrary to the collective interest of the organization)"* or *"Non-members of a project should not access project-internal files."*[3] The process of coordination is thus solely aimed at the *identification* of the most advantageous state of member behavior and says nothing about how this state is or could

[1] "Access" shall be understood in a broad sense here, not only including the reading of files but also writing or modifying them or even accessing other resources than files. Of course, the desired outcome could very well discriminate between these different activities. The optimal outcome could, for example, be reached by a certain member reading but not modifying a specific file. Such distinctions are discussed in more detail in section 6.2.

[2] The allocation of physical resources, which is usually also subject to the coordination process, is not considered further herein. We will concentrate on the behavioral part of coordination.

[3] Note that the statement is *"[...] should not access [...]"* and not *"[...] should not be able to access [...]"*.

be reached. The latter is subject to the task of *motivation*.[4] Large portions of the task of coordination are thus about what is usually referred to as *"risk management"*: assessing the different risks being induced by a certain kind of behavior (using WLAN, for instance), weighting these risks against the possible benefits, and deciding whether the considered behavior would be advantageous or not.[5]

Even without the consideration of motivational aspects, this task alone is – at least under today's circumstances – anything but trivial and trying to solve it will typically entail the usual costs of hierarchical coordination. These arise, as outlined in section 3.2.1, from

- a general inability of the management to gather and incorporate *all* relevant information into decision making (efficiency losses) and
- a general propensity to "overmanage" conduct inside organizations and a tendency to make decisions not only on the basis of economic but also of political and strategic considerations (bureaucracy costs).

Both of these cost types are, as we will see, also subject to the hierarchical coordination of information security inside organizations and will thus be considered separately. As we will also see below, the repeated changes of prevailing computing paradigms over time had significant impact on these costs. In order to develop a well-founded abstract understanding of security-related coordination and the associated costs, we will thus reconsider the three eras identified in section 2.1.4 in the light of efficiency losses of hierarchical coordination in sections 6.1 to 6.3. Reflections on bureaucracy costs will follow in section 6.4.

6.1 Efficiency Losses and Isolated Systems

In the original world of isolated systems represented by the first era of information security (see section 2.1.4), determination of the state of member behavior to be aimed for was comparably uncomplicated. Computers were handled by a small group of operators and no one except these operators had a need to access or otherwise interact with them. Due to the serial, non-parallel way of processing different jobs and due to the non-existence of shared resources, no interdependencies between different processes had to be considered. The desired state of member behavior was thus comparably easy to identify: Operators should access, modify, etc. the system whenever they need to.

[4] Planning which countermeasures should be put into place and how these should be used is also part of the *motivation* process. However, there is no doubt that motivation also has to be planned – large portions of the P from the above-mentioned PDCA-cycle refer to this subject, for example – but this would lead the discussion to another intellectual layer where one also had to think about how "motivators are motivated to perform in the way that was determined during the coordination of motivative measures", for example. We will ignore this aspect to simplify matters and to make the general point more clear.

[5] Notably, there is a growing body of literature suggesting that information security management should basically be regarded as risk management. See, for example, Blakley et al. (2001), Mercuri (2003), or Peltier (2005). Pinder (2006, p. 38) also raises the question whether information security should possibly *"be incorporated into the Risk Function"*.

Others, including non-members, should not (see figure 6.1). The information that was required for these considerations merely referred to the question whether a certain member had to be considered as being an operator or not. This information was – of course – completely available at zero or marginal cost and consequently, *all* relevant information could easily be gathered and incorporated into the coordination process. The considerations having to be made for the coordination of member behavior were thus of low complexity during the first era and the necessary information can be assumed to have been completely available.

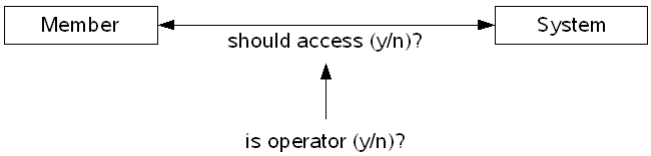


Figure 6.1: Coordination in era 1 – Member-system-relations

Furthermore, the number of decisions having to be made for coordination was comparably small. For every single member, exactly *one* decision had to be made with regard to any single system being considered: Should this member access the system or not? The maximum count of factors having to be included in the coordination process was thus the number of members multiplied by the number of considered systems. And finally, the decisions were comparably constant over time. Once it had been determined that the desired behavior for member A was *not* to access system X, this conclusion was rather a permanent than a temporary one.

The considerations having to be made for the coordination of individual member-behavior during the first era of isolated systems were thus of low complexity, made on the basis of relevant information being completely available, comparably small in overall count, and comparably constant over time. Resulting from these facts, only very few efficiency losses had to be expected for the coordination process during the first era. Neither had the central instance to deal with incomplete information leading to suboptimal outcomes of coordination nor did the quantity and complexity of required decisions exceed the capacity-limits of the central instance and make the introduction of additional hierarchical levels necessary. Consequently, coordination of member-behavior was possible at comparably low costs and the hazard of wrong decisions being made – the risk of *maladaption* – was virtually non-existent (see table 6.1).

Conclusion:
For the hierarchical process of coordinating security-related behavior during the first era of isolated systems, no material efficiency losses had to be expected. The number of relevant relations was small and each of the respective decisions could be made under virtually perfect information. This resulted in an *almost perfect outcome* of coordination during the first era.

Table 6.1: Efficiency losses of hierarchical coordination – Isolated systems (era 1)

	Era 1 (isolated systems)
Number of necessary decisions	low
Need for repeated reconsideration	low
Complexity of each decision	low
Level of uncertainty / lack of information	low
Expected level of efficiency loss	*low*

6.2 Efficiency Losses in the Mainframe Era

With the advent of terminal-based, shared systems in the second era, the coordination process had to be altered materially because of two significant changes. First, the sharing of resources like main memory or persistent storage by different processes led to the general possibility of *interdependencies*. One process could, for example, read, change or delete data that was actually belonging to another one. And second, the whole new paradigm of mainframe-based computing rested upon the principle of "ordinary members" accessing the systems themselves, thereby dramatically increasing their flexibility or, more abstract, their *possibilities of action*.

This principle of direct, flexible user access together with the general possibility of interdependencies rendered the established approach to coordinating member behavior increasingly useless: Had individual behavior still simply been considered as either "accessing" or "not accessing" a system, the process of coordination would possibly have come to the conclusion that a wide variety of members "should" and – as compared to the era of isolated systems – significantly fewer members "should not" access the considered system.

But if a multitude of users accessed a certain system, the possibility of interdependencies would have become increasingly relevant as any user would, for example, have been able to read, alter or delete data of others – be it by accident or even intentionally. Any member accessing the system would thus have imposed negative externalities in the form of heightened risk on all other members using the same system.[6] Without going more into detail, each additional user accessing the system would ultimately have decreased its overall security. The overall risks resulting from such interdependencies would in many cases have exceeded the overall benefits provided by a certain member accessing the system and consequently, the organization would be worse off than it had been without individual members accessing the system. Even if

[6] See again the respective considerations on negative security externalities in section 4.1.1.

considerable benefits could be derived from "ordinary members" accessing the system directly, these benefits would at the same time have been nullified or exceeded by the risks arising from interdependencies.

The alternative outcome of the coordination process could thus have been that only for very few members the overall benefit of accessing the system would have been considered as outweighing the resulting risks and that only those members "should access" the considered system while all others "should not". This would have countervailed the risks arising from resources being shared among (too many) different users but would in turn have decreased realizable benefits, too. In principle, the outcome of coordination would have been nearly the same as in the first era: Only very few members "should access" the system, others "should not". Most of the possible gains from technological development from isolated to shared systems had in this case been passed up.

Obviously, neither the substantial risks having to be taken for the case of giving access to a multitude of members nor the abdication of substantial gains from technological development would have been in the interest of organizations. The challenge for information security was thus to minimize the risks arising from the newly existing interdependencies while at the same time still allowing to realize the gains from the shared use of systems.

The answer to this problem was to consider individual member-behavior in more detail than before. Instead of coordinating activities in respect of whole systems, considerations were now made in relation to single information resources such as files, database tables, etc. This heightened the granularity of the coordination process and allowed for more sophisticated outcomes of the coordination process, thereby minimizing the negative effects mentioned above. A certain member that would have been considered to "should access" a certain system in order to perform a certain job without this increased granularity could now be determined to "should access" only those specific files that are actually required, for example.[7]

Additionally, these accesses were also differentiated by their mode. In the first era, the set of possible "behavioral patterns" being considered at all consisted of exactly two elements: "access" a certain system and "do not access" it. In the second era, this set got more diversified, now distinguishing different activities like reading, writing, changing, or appending[8] data, increasing the detailedness of the coordination process even further. Overall, member behavior was thus contemplated at significantly higher granularity in the mainframe era than in the era of isolated systems to allow for more sophisticated allocations of activities and to thereby counteract the possible drawbacks

[7] Of course, there was still a need for some operators to physically access the systems that had to be considered during the coordination process, too. However, the respective decisions could still be made in the established manner without significant efficiency losses and their relevance decreased continuously in comparison to the relations described here. The still existing member-system-relations will thus be put aside to illuminate the more decisive aspect of "ordinary members" accessing the system in more lucidity.

[8] Hoffman (1977, p. 25), for example, mentions read, write, execute, delete and append. Anderson (2008, pp. 96 ff) initially distinguishes between read, write and execute but later also mentions activities like taking ownership, changing permissions, and deletion (p. 102) to be considered separately.

of the shared systems approach.

For the costs of coordination, these changes had implications, of course. Most decisive here is the significant rise in the number of relevant relations. Instead of coordinating a comparably small amount of relations between members and systems, the process of coordination now had to determine *which* member should perform *what kind* of access to *which* resources in order to reach the most efficient outcome for the whole organization. The overall count of relations being relevant for coordination was thus the number of members multiplied by the number of resources (files, database tables etc.) multiplied by the number of possible behavioral patterns (read, write, append, or delete data, for example). Determining an overall desired state of member behavior that optimizes the overall outcome of the organization – that is, implementing "perfect" coordination – would have required *all* of these relations to be examined explicitly.[9] A further increase in the number of relations having to be considered results from the need for repeated reconsiderations in response to members changing functions, for example, where any such switch basically had to lead to a reconsideration of all relations regarding the specific member.[10] Compared to the rather fixed distinction of operators and non-operators from the first era, this also resulted in additional considerations of relations having to be made during the coordination process.

Furthermore, each of these relations was more complex to estimate than it was the case during the first era. Instead of simply having to distinguish operators from non-operators (of a certain system, maybe), it basically had to be considered what value would be generated if a certain member accessed a certain resource in a specified manner and what risks would result from this specific access. Especially because of the general possibility of interdependencies – or, externalities – this valuation could get considerably complex. For instance, the overall value arising from member A changing the content of a certain file X containing contact data of all members would typically depend on the value that other members attribute to the accuracy and integrity of that file. If no other member used such a file, the outcome of a "perfect" coordination process would presumably differ from a situation with thousands of other members regularly using it. Thus, not only did the number of relations having to be considered during the coordination process rise significantly in the second era but also did the complexity of estimating every single relation increase because of more factors having to be taken into account.

As a final aspect, the availability of the necessary information for each of these considerably complex estimations is also relevant for the costs arising from efficiency losses in the hierarchical process of coordination. In the first era, this information had been limited to the question whether a certain member had to be treated as operator

[9] See especially Ware (1970, p. 12): *"The combination of a clearance and a need-to-know constitutes the* necessary and sufficient *conditions for granting access to classified information"* (emph. added). In this concept, access may basically not be granted without *explicit* consideration of any single relation between a user and a resource of information.

[10] Of course, such changes could also apply to information resources, not only to the members. See again Ware (1970, p. 10) mentioning possible *"changes in the classification and sensitivity of the files"* being relevant, too.

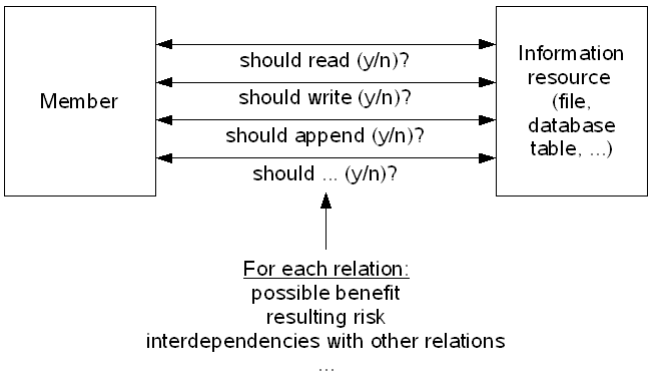


Figure 6.2: Coordination in era 2 – Member-resource-relations

or not and the information could thus be considered as being completely available to the central instance (see above). For the second era, the information having to be factored into the coordination process was more comprehensive, including details like the specific functions of individual members, their assignment to and specific roles in different projects, or the possible impact that could be impaired upon them by other members provoking different kinds of negative externalities. All this and much more information had to be taken into account and had thus to be available to the central instance to accomplish "perfect" coordination. It can be assumed that this state of perfectly available information will usually not have been given and that it could – if at all – only be reached at considerable cost. The central instance thus had to realize coordination under conditions of incomplete information and thus with a certain amount of uncertainty.

Given these facts, we can now derive a first basic model of efficiency losses having to be expected for the hierarchical coordination of security-related behavior during the second era. As mentioned above, the number of relations for which a desired outcome had to be determined increased significantly and each of these determinations additionally necessitated – in comparison to the first era – more complex considerations. From the perspective of coordination costs, this gives rise to the problem of the central instance reaching its capacity limits of gathering and processing the respective information.[11] It can be expected that the central instance was not able to precisely determine the desired outcome for every single relation anymore and that the respective decisions thus had to be made without taking all relevant factors into account.[12] In a very basic model of coordination during the second era, decisions being made

[11] Even if from a more pragmatic perspective, the problem associated with possible capacity limits of a central instance was also recognized by Saltzer (1974, p. 389): *"If a system design forces too many administrative decisions [...] to be set by a single administrator, that administrator can quickly become a bottleneck and an impediment to effective use of the system [...]."* Regarding the effect of increased complexity, see also Bailey (1993, p. 3), concluding that *"any security policy or management scheme based on 'global understanding' of the network is bankrupt"*. The effect he describes perfectly resembles the notion of strongly increased efficiency losses pursued herein.

[12] Remember the notes on the non-measurability of information security's value given in section 4.1.3.

by the central instance would thus be based on a certain amount of incomplete or inaccurate information, leading to the possibility of wrong decisions being made and thereby resulting in efficiency losses (see table 6.2).

Table 6.2: Efficiency losses of hierarchical coordination in era 2 – Basic model

	Era 2 (centralized systems)
Number of necessary decisions	medium – high
Need for repeated reconsideration	medium
Complexity of each decision	medium – high
Level of uncertainty / lack of information	medium
Expected level of efficiency loss	*medium – high*

Even if considering such a central instance that explicitly decides on every single relation between members and resources for the whole organization is highly worthwhile for illuminating the underlying economic principles, it does obviously not represent actual practices of information security during the second era. Instead, organizations used, and nowadays still use, different well-known and widely established mechanisms like local responsibilities, role- or group-based access, or default permissions to realize organization-internal information security. As we will see in the following, these approaches can – from an abstract point of view – reduce efficiency losses having to be borne in comparison to the basic model outlined above. They thus serve the organization's overall goal of optimizing the coordination process.

6.2.1 Optimization

As outlined above, the increased number and complexity of relations having to be considered would give rise to capacity limits of the central instance being reached and to decisions being made upon the basis of necessarily insufficient information. To overcome the resulting efficiency losses, economic theory as outlined in section 3.2.1 suggests decisions to be delegated to lower hierarchy layers by the central instance.

For the field of organization-internal information security, this delegation could, for instance, be realized by departmental security officers. A separate security officer of department A could be responsible for determining the desired outcome of security-related behavior for all members of that department. Project managers could decide over the behavior of project members or over the desired behavior of all members in relation to the respective project files.[13] Any such approach would disburden the central instance substantially and each decision could be based on more complete information. In the outcome, the overall loss would be decreased.

[13] Saltzer's (1974) concept of *"project administrators"*, for example, exactly falls into this category.

However, there would still be some interdependencies between the individual behavior of members from department A and members from department B that also had to be factored into the coordination process. To include these interdependencies, the respective departmental security officers had – due to existing information asymmetries between these two departments – to communicate with each other to adjust the respective sub-processes of coordination. As outlined in section 3.2.1, this communication – be it performed directly or through the central instance – had to be expected to be subject to a certain amount of content losses and alterations and thus to *control loss*.[14] Even if delegation could thus overcome the problem of capacity limits of the central instance and thereby decrease efficiency losses, it would in turn result in yet another kind of efficiency losses. But these can very well be lower than the original losses resulting from insufficient information. Especially in environments with few interdependencies between different departments having to be considered, the mechanism of delegation could thus be used to reduce the overall losses being present in the hierarchical coordination process.

Besides delegating decisions to lower layers, coordination costs were also countervailed by *generalizing* decisions. As one of the main reasons for capacity limits being reached was the significant rise in relations that had to be considered and decisions that had to be made, a reduction of these numbers could also reduce overall efficiency losses. By generalization, the central instance could merge different relations into a single one and consider this generalized relation more consciously and on the basis of more substantial information. Such generalizations can be used in various forms, but basically, it is possible either to generalize over members or to generalize over resources.[15]

The most obvious generalization of members is the use of roles instead of considering single individuals. Basically, a role can be seen as a virtual individual that can be adopted by different real members but is considered as single entity during the coordination process.[16] For example, it can be determined during the coordination process that the respective receptionist "should access" the organization-internal contact database whatever real member currently holds this role. In this case, it had only to be considered once which activities the respective receptionists should and which ones they should not perform.

A comparable but slightly different form of generalization is to consider groups in-

[14] Alternatively, one could also ignore the existence of interdependencies between different departments. Nonetheless, this would also lead to certain losses in the coordination process.

[15] From a more pragmatic and more technology-oriented perspective, Anderson (2008, p. 97) points into the same direction, stating that *"[w]e will usually need a more compact way of storing and managing [the information about member-resource-relations]. The two main ways of doing this are to* compress the users *and to* compress the rights*"* (emph. added). Hoffman (1977, p. 28) additionally mentions the generalization over terminals. Access from different terminals is not considered explicitly here but would otherwise have increased the number of relations even further. In this case, the respective generalization would represent another way for minimizing the number of relations.

[16] For a seminal introduction to role-based access control, see Sandhu, Coyne, Feinstein, and Youman (1996).

stead of single members.[17] In this case, multiple members can be treated collectively in their relation to certain information resources. For example, all members of a project can – independently from their role in that project – be determined to "should access" a certain project-related file. Again, the relations would not have to be examined separately. Without going more into the details of the two concepts[18], roles as well as groups thus allowed to reduce the overall number of necessary considerations significantly and thus helped to reduce the adverse effects described above – even if there were always some members that did not fit into one of these roles or groups and had to be considered individually.

Besides merging members, the number of relations could also be reduced by merging information resources together and handle them as single entities. For example, all files regarding to a certain project might be considered like a single entity with the coordination process only trying to determine whether a certain member should access those files or not. The same kind of generalization is done when system- and user-files are distinguished or when data is attributed to a certain department and treated conjointly during the coordination process. There exists a multitude of different models for such a generalization over information resources[19] and all of them serve the goal of reducing the number of relations having to be considered explicitly.

Of course, both kinds of generalization could also be combined to reduce the amount of relations having to be considered even further. For example, one could only consider the relation between the generalized group of members working in a certain project and the generalized resource of all files belonging to the project, consider the relation between receptionists and "reception-related files" or determine how ordinary members of the billing department should behave in relation to customer data.[20] Again, every such generalization reduces the number of relations having to be considered.

On the other hand, such generalized considerations necessarily hide specific aspects of the original and more detailed relations. By generalization, the higher degree of granularity originally introduced for shared systems gets *de*creased again to a certain extent and this entails the general risk of misalignment. Not every member belonging to a project might, for example, have been determined to behave in the same manner with regard to single project files if all relations were examined separately. Perhaps, only one or two members "should write" into the respective project schedule while others "should read" only in the optimal case. While generalization allows to reduce the number of relations and thereby the efficiency losses of hierarchical coordination,

[17] Some authors of that time, like Friedman (1970, pp. 269 f), called such groups of users *"cliques"* while referring to groups in the case of generalization over data or information resources. Obviously, terms were not always as clear as nowadays.

[18] For a more extensive distinction between the concepts of roles and groups, see, for example, Anderson (2008, p. 98).

[19] See, for example, the approach of *"compartmentalization"* of different items of data suggested by Friedman (1970, pp. 268 ff). Security models using confidentiality or integrity levels like Bell and La Padula (1976) or Biba (1977) can also be interpreted as generalizing over information resources and thereby limiting the number of considered relations.

[20] Another kind of combination is that described by Bell and La Padula (1976, pp. 13 f), where the generalization by a classification-level and the generalization by a formal category are combined and furthermore applied to information resources as well as to members.

it also entails losses – and thus costs – due to some specific aspects being ignored. These costs can be referred to as the *costs of over-generalization*.

Generalization thus offers the opportunity to reduce costs arising from capacity limits but at the same time entails a certain amount of costs resulting from non-consideration of specific aspects. The less generalization is present, the higher will be the costs resulting from capacity-limits and inadequate information and the lower will be the costs arising from over-generalization. And correspondingly, the more generalization is present, the lower will be the effect of capacity-limits and the more considerable will be the costs of over-generalization. The efficiency losses resulting from capacity-limits and from over-generalization can thus be represented graphically as done in figure 6.3.

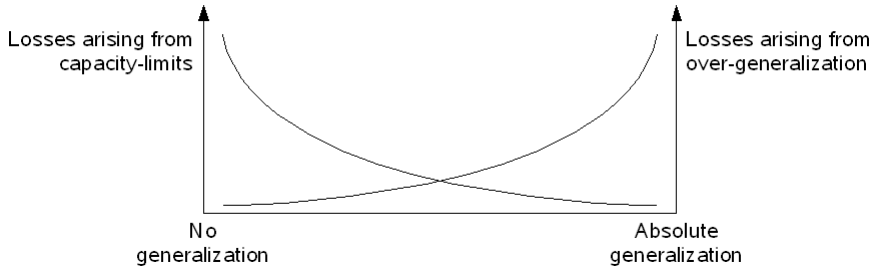


Figure 6.3: Efficiency losses resulting from capacity-limits and optimizing generalization

For the extreme case of no generalization being present, we would result in the above case of any single relation between any member and any information resource having to be considered separately – perhaps optimized by some decisions being delegated to lower hierarchy layers. As mentioned above, this was obviously not what happened during the second era. The other extreme would be to treat all entities in the same manner. This might be hard to imagine for users, but for information resources, the effective outcome of such an absolute generalization is quite clear: It would result in exactly the same cost structures that prevailed during the first era. Different information resources and different kinds of access would not be considered separately anymore but rather as merged entity which the member either "should access" or "should not access" as a whole. For reasons outlined above in more detail, this would also result in undesired outcomes.

A special case of generalization is the use of defaults. Instead of examining every single resource in relation to every single member, default decisions about member-behavior with regard a certain resource could be made implicitly and without conscious considerations. The initial access control lists being used in MULTICS environments to *"minimize the explicit attention"* (Saltzer 1974, p. 391) are an example for such implicit default decisions. Of course, the same approach can also be applied to members by providing them with some initial access rights that implicitly represent some default decisions about their desired behavior. Like other generalizations, defaults

can minimize the number of necessary considerations but carry along with them the inherent risk of being inadequate in certain cases.

Altogether, there are several mechanisms for reducing the efficiency losses that result from hierarchical coordination of security-related member behavior during the second era of centralized mainframe computers. By (partially) delegating the coordination process to lower hierarchy layers, the capacity-limits of the central instance can be counteracted for the price of a certain control-loss arising. Generalization in its various forms outlined above also allows to reduce costs arising from control loss by limiting the number of relations having to be considered, thereby allowing each remaining relation to be examined more consciously and on the basis of more information. On the other hand, generalization also holds the risk of over-generalization resulting in too less differentiation between different relations. And finally, the same is also true for the use of defaults: The number of relations having to be considered is reduced but a certain risk of making inadequate decisions arises.

For any of these optimization approaches, the challenge thus consists in finding the optimal intermediate state that minimizes the overall efficiency losses of hierarchical coordination. Naturally, this is what was done in practice during the second era. Some relations were treated by the use of a certain number of abstract roles and some not. Some information resources were treated collectively while others were still considered as singular entities. And default permissions were used for some cases and for others not.

Table 6.3: Efficiency losses of hierarchical coordination – Centralized systems

	Era 2 (centralized systems)
Number of necessary decisions	medium
Need for repeated reconsideration	medium
Complexity of each decision	medium
Level of uncertainty / lack of information	low – medium
Additional drawbacks due to optimization	low – medium
Expected level of efficiency loss	*(low –) medium*

By all those means, it was possible to substantially reduce the overall efficiency losses of hierarchical coordination as compared to the basic model of every relation having to be considered separately that was developed above. The number of relations could be lessened to a *medium* level and the level of uncertainty under which each decision had to be made could also be alleviated from a medium to a *low to medium* level. In return to these reductions of efficiency losses, some additional costs had to be borne as results of control-loss and of a certain over-generalization being inevitable. But if used in a deliberate manner, these mechanisms allowed to reduce overall efficiency

losses of hierarchical coordination. Altogether, this resulted in a *(low to) medium* level of efficiency losses having to be accepted by organizations for the hierarchical coordination of security-related member behavior during the second era of centralized, mainframe-based computer usage (see table 6.3).

Conclusion:
The advantages offered by the shared systems of the second era could only be utilized with "ordinary" members accessing systems directly and with their individual possibilities of action being broadened. This necessitated considerations about security-related member behavior to be made with higher granularity and thereby significantly increased the overall number of relations that had to be considered during the coordination process. This, in turn, led to the risk of the central instance reaching its capacity limits and of inadequate decisions being made. The resulting efficiency losses could be limited through different approaches of optimization (delegation, generalization, defaults) which either decreased the number of relations that had to be considered or lowered the level of information that was factored into considerations. Even if these optimization approaches also entailed other losses, they allowed, if used deliberately, to limit overall losses to a *(low to) medium* level

6.3 Efficiency Losses in the PC Era

As delineated above, the switch from isolated to shared systems led to a substantial expansion of members' possibilities of action and caused substantial changes for the coordination of security-related member behavior inside organizations. With the advent and establishment of Personal Computers (PCs)[21] and the replacement of centralized, mainframe-based systems, this development continued even further. Again, the changes were manifold and shall only be reflected in a generalized manner here to illuminate the underlying abstract principles and problems as clear as possible. Like it was done for the case of mainframe-based computing above, we start our considerations with the hypothetical scenario of no optimizing mechanisms being in place.

From such an abstract perspective, two main changes characterize the shift from mainframe- to PC-based computing. First, members now basically had unlimited access to and control over the actual computing resources. Instead of having to act within the well-defined boundaries of a centrally controlled mainframe environment, they were now able to perform *any* kind of access to *any* component of the system being used. They could execute self-written or self-supplied programs, change any aspect of the system's configuration or even alter the operating system being executed.[22] Furthermore, the basically unlimited access also allowed members to change the physical setting by attaching additional devices like printers, storage devices or, later,

[21] The term "PC" is used in a broad sense here, not only including IBM-compatible personal computers.

[22] See, for example, Murray (1986, p. 7): *"On the personal computer, anybody can invoke the operating system. All I have to do is walk up with my diskette and I can load my own operating system."*

modems.[23] In principle, they could even change internal components like processors or fixed storage media.

Through this absolute local power, members were able to define their productive working environment by themselves according to their personal needs, allowing them to use "their" systems in any manner they desired. In the more abstract terminology already used above, the establishment of PCs thus broadened the members' possibilities of action even further. Even if it was this local power exerted by "ordinary" members and the resulting possibility to adapt the working environment to specific requirements that *"spurred ingenuity [and] increased personal productivity"* (Carr 2005, p. 69), it also induced, as we will see later, substantial changes for the coordination as well as for the motivation process.

The second characterizing change regards the role of central instances. Within PC-based environments, there was – generally speaking – no central point for exerting full control over member-resource-relations as introduced in section 6.2 anymore. As personal computers typically possessed local storage capabilities in the form of hard-disks or removable media, members were able to create, change or delete information resources such as files or local databases, to exchange data with each other by using floppy disks and to ex- and import data from and to the systems by themselves without any central instance being involved or even being able to notice those activities. Different from the former mainframe-based environments and without any further mechanisms being in place, PCs thus brought with them the inherent property of decentralized storing and processing, thereby contradicting the formerly established approach of a dedicated single point allowing to oversee the entirety of all member-resource-relations.[24]

In a first stage of development, the implications with regard to security-related member behavior were comparably simple. PCs were not interconnected with each other but rather operated in an isolated manner. No noteworthy kind of resource-sharing was applied and consequently, there were no system-immanent possibilities for negative externalities in the form of different members' processes interfering with each other.[25] In a way, PCs thus represented yet another incarnation of the isolated

[23] In fact, the more recent problems regarding "rogue access points" (see, for example, Godber and Dasgupta 2003) or various types of USB-devices are also specific manifestations of the same abstract problem.

[24] See also Steinauer (1985, pp. 1-5 ff) stating that *"[o]ne of the perceived benefits of personal computers is the reduction of users' dependence on (and, perhaps, frustration with) a central data processing facility."* With a specific focus on information security, Steinauer considers physical accessibility, a lack of built-in security mechanisms, the person-related and more sensitive nature of data being handled, and the responsibility of users instead of specialized personnel as the most relevant specifics of PCs in comparison with the former mainframe systems (pp. 1-2 ff). See also Quillard, Rockart, Wilde, Vernon, and Mock (1983, pp. 50 f) discussing fundamental differences between mainframe-based and PC-based environments within the organizational context.

[25] Actually, there was a certain possibility of negative externalities arising as a consequence of different members using the same system successively with the same storage media, for example. Furthermore, important data could also be given on floppy disks from one member to another and, intentionally or by mistake, altered or destroyed by the recipient (see Steinauer 1985, pp. 1-5 f). Nonetheless, such aspects played a minor role and we will consider the more interesting case of substantial externalities being possible below, anyhow.

systems from the first era – even if these "isolated systems" were now accessed by "ordinary" members.[26] Due to the insignificance of externalities, the coordination of security-related behavior could thus be based on member-system-relations again, requiring far less relations to be considered than the coordination process for the shared systems of the second era did. Generally speaking, considerations could – again – be reduced to the question who "should" access a certain system and who "should not".

But most important and anything but trivial was the problem of identifying *all* relevant systems.[27] As mentioned above, PC-based computing was possible without any centralized instance being involved and this was not only the case for the access to information resources but also for the operation of PCs in general. Consequently, members could use computers without any central instance even being aware of the systems' existence. Even if the coordination process of security-related behavior could in this first stage of PC-usage be realized on the basis of member-system-relations again, certain losses had to be expected as a result of central management's inability to identify all relevant systems.

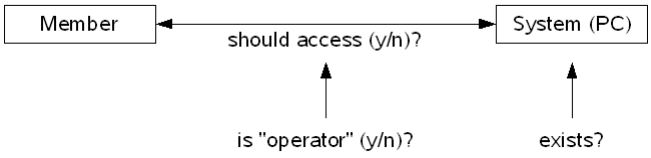


Figure 6.4: Coordination in first stage of era 3 – Member-system-relations

Nonetheless, like there had been benefits from the shared use of resources during the second era, these benefits naturally also apply with PCs being used instead of simple remote terminals. The shared use of specialized hardware, the possibilities of electronic communication or the sharing of information among different members through databases and file repositories still represented mechanisms for realizing considerable benefits for the organization as a whole. It was thus only consequential that organizations began to interconnect the initially isolated PCs to realize those possible gains.

The resulting organizational IT infrastructures represented some kind of hybrid between the initial state of isolated PC usage and the centralized paradigm from the

[26] See also Panagacos, Farber, Dreyer, and McVeigh (1985, p.): *"By having the users introduce these machines in a vacuum is causing a situation environment analogous to the centralization seen in the 70's."*

[27] In fact, this problem was widely discussed during a certain period. See, for example, Isaacson, Osborne, Gammill, Tesler, Heiser, and Warren (1978, p. 94), predicting that *"[t]he corporate data processing center will lose control of the data processing function [...]."* Later in 1983, Rockart and Flannery observed exactly this in a field study of end user computing: *"It was recognized in each company that the Information Systems department could* not *control the use of end-user computing resource"* (emph. in original). See also Benson (1983), vividly describing how poorly IS managers were informed about the actual use of PCs within their organizations at all and Quillard et al. (1983, p. 51) noting that *"users are now able to purchase their own hardware and software, and are doing so eagerly in the companies [...] visited."*

second era. The first approaches to combine the advantages of decentralized PCs with those of centralized mainframe systems consisted in the employment of terminal emulations which allowed to use PCs as terminals to access existing mainframe systems. However, with the earliest of those terminal emulations PCs could only be operated either as a PC or as a terminal at a given time.[28] Even if this model allowed to use capabilities from both worlds and to enhance functionalities available to the user, it did not exploit the full potential that would have been possible through integration of the two modes of operation. This integration was what happened next.

The easiest example for the simultaneous and integrated use of decentralized PCs and centralized instances are file repositories shared among PC users to support collaboration. Like it had been the case during the mainframe era, different members could access the same information resource to generate value for the whole organization from doing so. But in contrast to mainframe-based environments, members now could use those information resources locally on "their" PCs – with all the increased freedom and flexibility but also with the unlimited access and the limited overseeability that constitute the paradigm of decentralized PC usage. Members were, for example, able to perform reading access to a centrally stored information resource, copy it to a local medium and ultimately exert full control over this local copy. From an abstract point of view, the model of decentralized, interconnected use of PCs thus allowed members to move information resources out of the scope and coverage of centralized instances and carry it over to a mode of unlimited local access under the PC-specific conditions of extensive possibilities of action. Other information resources might even never have reached the scope of any centralized instance at all.[29]

Now, let us consider the implications for the process of coordinating security-related member behavior within such environments.

First, and like it was the case for the introduction of mainframes, the fact of different members sharing the same resources led tho the possibility of negative externalities. As members were not strictly separated from each other but rather accessed the same resources conjointly, any member could basically behave in a manner that imposes damage to the others in the form of deleted, altered or publicized information from the central instance, for example. Due to this risk being added by any single member accessing the central, shared system, the considerations that were made for mainframe systems with regard to the non-suitability of simple member-system-relations[30] apply here, too. Like it was the case for the second era, member-resource-relations had thus to be considered for centralized information instead (see figure 6.5).

Nonetheless, the coordination process for PC-based environments differed from the former mainframe-based case in various respects. In particular, PCs gave rise to

[28] See, for example, Murray (1984, p. 301): *"[T]he software permits the personal computer to function as a terminal or as a computer, but not as both at the same time."*

[29] See, for example, Benson (1983, pp. 43 f), mentioning that much of the information being handled within local PC environments was *"so local in nature that it [was] not cost-effective to centralize it."* The possibilities arising from mechanisms like the sharing of local folders are not even considered explicitly here. Generally speaking, these increased the problem of information resources being located out of the scope and coverage of centralized systems even further.

[30] See page 119.

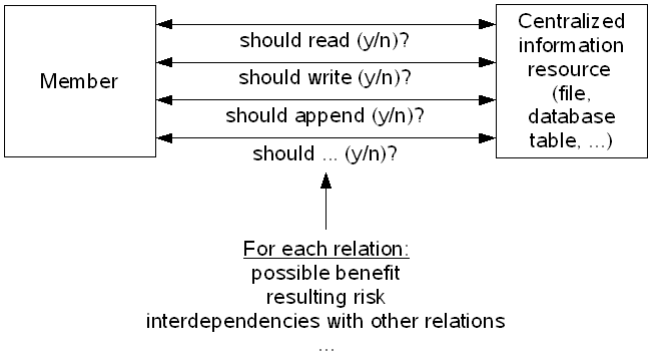


Figure 6.5: Coordination with regard to centralized information in era 3 – Member-resource-relations

completely different kinds of information security risks and, resulting from these, to completely different kinds of negative externalities being possible than before.

Even if a member himself could, for example, have had a certain benefit from connecting a modem to his PC to access the system remotely, such a modem would also have induced considerable risks to the organization as a whole as attackers could possibly use it as a gateway to the complete internal network. Another example are external devices being used to store personal data or to take work home over the weekend. Doing so could very well represent a benefit to the individual member but it could also, emanating from the member's PC, possibly result in the infection of the whole organizational network with different kinds of malware and thereby also impose negative externalities. And finally, the local installation of a specific software or even of an alternative operating system could constitute considerable benefit for an individual member but on the other hand corrupt the security of the overall environment to a certain extent, too.

Many further examples could be given but what unifies the mentioned cases is that they all go far beyond the former model of simply regarding member behavior as performing a specified mode of access to a known and well-defined information resource within a well-defined environment. In all cases, there are entities – technical components or information resources – that can, as a matter of principle, not be known completely by a central instance and there is member behavior that can, also as a matter of principle, not be foreseen in entirety. If the possibility of a member attaching a modem to a PC is not anticipated at all, this behavioral pattern can, of course, hardly be considered during the coordination process. Abstractly speaking, member behavior had thus to be coordinated in relation to a basically *undefined* set of technical as well as information resources and, furthermore, with regard to an also *undefined* set of possible behavioral patterns.[31] And of course, interdependencies

[31] Of course, it was not only the use of PCs instead of terminals that caused these uncertainties. The generally increased complexity of systems, the growing importance of the Internet, the ever-increasing number of information resources being handled in general, the establishment of private

between the different relations were still possible and had to be considered, too (see figure 6.6).

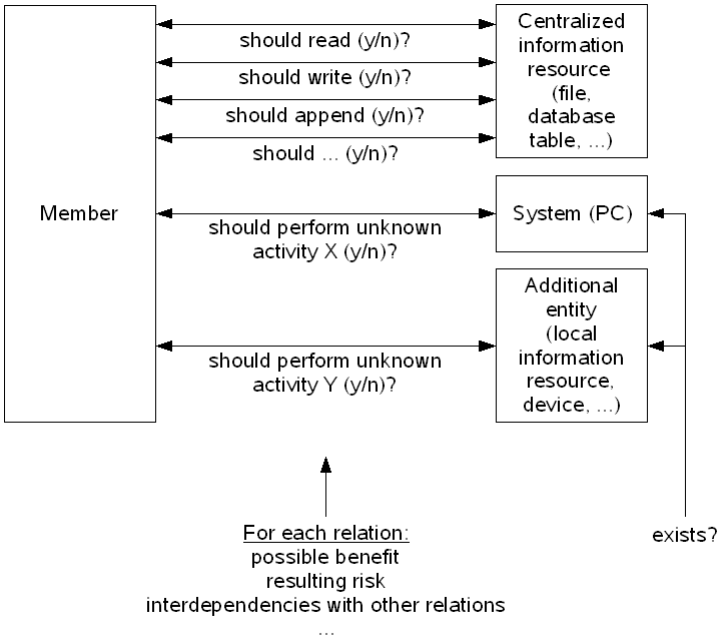


Figure 6.6: Coordination in era 3 – unknown relations and behavioral patterns

In this original, idealized model of PC usage without any further optimization mechanism being in place, the efficiency losses of hierarchical coordination would necessarily have been extensive. Even if the number of information resources being shared through the still-existing central instance is assumed to have been constant, the overall number of relations that had to be considered in order to realize "perfect coordination" increased because of the newly accrued relations between members on the one and local systems and additional resources on the other hand. Independently from the general problem of knowing about the existence and the quality of these relations at all, they had basically to be included in a coordination process that should achieve perfect coordination of security-related member behavior. But once a certain relation had been considered, the need for repeated reconsideration did not differ much from the second era. Member functions might have changed as well as the quality of existing information resources from time to time but not substantially more often than it had been the case in mainframe-based environments. And once one of the new relations – between a member and a certain device, for example – had been examined, there

use of IT and many other factors surely played significant roles, too. But all these developments led into the same direction of significantly increased uncertainties that posed the major challenge to the coordination process during the third era. See also Anderson (2008, p. 96) referring to the former world of mainframe-based computing as *"the lost Eden of order and control"* in contrast to *"the messy reality of today"*.

would in most cases have been no reason to assume that this relation would have needed frequent reconsiderations.

Contrary to the need for repeated reconsideration, the complexity of considering any single relation did very well increase in comparison to the basic model of the mainframe era. As the complexity of the overall system increased as well as the number of other relations also having to be taken into account, the overall benefit of a certain behavior depended on even more factors than it had been the case for the second era. Especially the new relations that were not covered by a well-defined set of possible behavioral patterns required more sophisticated considerations. Deciding about the overall benefit and the overall risk of external storage media being used by a certain member was unquestionably a significantly more complex task than deciding over the possibility of reading the content of a certain and well-defined file. We can thus consider the complexity of each decision as being of a high to very high level.

But most significant for the third era was the strongly increased uncertainty under which the coordination process had to take place. As mentioned above, neither was the existence and nature of all relevant entities known to the central instance nor could the central instance know all kinds of behavior that a certain member could perform in relation to the diverse entities. Even if the increased autonomy and flexibility of the single members were the primary advantages of PCs in comparison to mainframe-based environments, they also led to substantial information asymmetries between the central instance and the different members. Due to these asymmetries and to the increased number and complexity of necessary decisions, considerable efficiency losses had to be expected for the hierarchical process of coordinating security-related member behavior within PC-based environments (see table 6.4).[32]

Table 6.4: Efficiency losses of hierarchical coordination in era 3 – basic model

	Era 3 (decentralized PCs)
Number of necessary decisions	high – very high
Need for repeated reconsideration	medium
Complexity of each decision	high – very high
Level of uncertainty / lack of information	very high
Expected level of efficiency loss	*high – very high*

Of course, this idealized basic model does again not represent how organizations

[32] Again, note the strong relation between these efficiency losses and the non-measurabilty of information security's value delineated in 4.1.3. Were this value measurable, the coordination process could, at least for the risk-part, rest upon such measurements and be more efficient. But for the reasons given here, an accurate measurement of risks is virtually impossible, leading to a strongly inefficient coordination process. There are thus economic reasons for the non-measurability of information security that we observe nowadays.

really implemented the coordination process for the PC-based environments of the third era. Had they actually tried to consider any single relation exhaustively, they would either have reached the capacity limits of the central coordination instance immediately, or they would have had to base decisions on unquestionably insufficient information. In both cases, organizations would have had to bear enormous costs arising from suboptimal decisions being made or, more abstractly, from efficiency losses. Like it had been the case for the coordination process of the second era, different approaches for optimization were thus developed over time.

6.3.1 Optimization

The simplest and most obvious way of counteracting the significant efficiency losses arising from the use of PCs instead of mainframes is, of course, to eliminate the specifics of PCs that cause the significant rise of inefficiencies. In the extreme case, this would have led to PCs simply being used as strongly overdimensioned terminals within de-facto mainframe environments and all benefits possibly arising from flexibility and user autonomy would have been relinquished. Even if this approach is actually pursued in a couple of cases nowadays, it does for most organizations not adequately address the specific requirements of the third era and will thus not be considered further.

Instead, mechanisms were needed that allow to preserve as much as possible of the advantages of PC-based in contrast to mainframe-based environments while at the same time minimizing efficiency losses as far as possible. As we will see in the following, most of the mechanisms that were used during the third era to minimize efficiency losses do not differ much from those of the second era on an abstract level. Basically, most of the approaches for the third era can also be subsumed under the fundamental concepts of *generalization* and *delegation*.[33]

Regarding generalization, we already identified the approach of generalizing over members through role- and group-based approaches and the generalization over known information resources in section 6.2.1. These approaches could of course also be applied to PC-based environments to reduce the number of relations having to be considered. To meet the specific requirements of the third era, generalization over users and information resources has also been translated to local systems to a certain extent. The well-known and widely established practice of generalizing over "system-related files" and "ordinary members" and coordinating behavior in relation to these generalized entities may serve as simple example. In this case, the coordination process only had to determine whether ordinary members "should" or "should not" access system-related files in a specific manner.

Besides reducing the overall quantity of relations, such generalization also allows to include those local information resources in the coordination process that are principally unknown to the central instance. This is also the case for the generalization over unknown *technical* entities. It is, for example, established practice to generalize over different and basically unknown kinds of locally used devices by considering abstract classes like "external devices with storage capabilities" and to realize coordination of

[33] The abstract concept of defaults will not be explicitly considered here because it represents – as already mentioned in section 6.2.1 – a special case of generalization.

security-related member behavior on the basis of this abstract class. Similarly, it is also possible to generalize over basically unknown systems by coordinating member behavior in relation to abstract classes of systems like Windows-based PCs, notebooks, or – as a border case between device and system – smartphones. Generally speaking, such generalization allows to reduce the number of relations having to be considered and thereby to counteract the problem of capacity limits.

As an additional specific challenge of the third era, we identified the problem of coordinating member behavior in relation to a basically undefined set of possible behavioral patterns. Generalization has been applied here, too. A classical example for generalizing over member behavior is the approach of considering all possibilities for changing a certain system's physical configuration conjointly. Instead of trying to examine any potential physical modification separately, organizations usually only ask whether a certain member "should" or "should not" change a certain system's physical setting. Like the generalization over information resources, devices and systems, this generalization over member behavior allows, on the one hand, to minimize the number of relations having to be considered and, on the other hand, to include unknown and unforeseen behavioral patterns in the coordination process.[34]

In all these cases, generalization allows to reduce the number of relations having to be considered and also addresses the problem of unknown entities or unknown behavioral patterns to a certain extent. Generalization thus again allows to reduce the inefficiencies of hierarchical coordination. Nonetheless, the effect of generalization always hiding some specific aspects, provoking too less differentiation between different relations and thus leading to the possibility of substantial misallocations outlined in section 6.2.1 applies here, too.

Besides the abstract concept of generalization, the *delegation* of decisions to lower layers was also mentioned as possible approach to reduce the inefficiencies of hierarchical coordination in section 6.2.1 as well as in section 3.2.1.

The concept of departmental or site-specific security officers being responsible for isolated domains of the coordination process was already discussed in section 6.2.1. This approach was retained during the third era with the same pros and cons. On the one hand, local security officers allow to partially overcome the problem of capacity limits and insufficient information of the central instance. In respect of the specific conditions of the third era, this is particularly relevant for lowering the uncertainty and lack of information about local circumstances that the central instance had to cope with otherwise.[35] On the other hand, possible interdependencies between different departments also necessitate communication between different local security officers and again raise the problem of content losses and alterations.[36]

[34] Again, all of these generalizations could also be combined, leading to relations between "members of department *A*" and "external devices with storage capabilities" or between "ordinary members" and the activity of "altering physical configurations" being considered during the coordination process, for example.

[35] This is especially the case when the concept of delegation is applied in depth through the nomination of members being responsible for information security on a workgroup level. These will, for example, face a significantly lower uncertainty about the use of unknown devices and the benefit arising from that use than the central instance.

[36] See also the BSI (2005a, p. 18), pointing out that *"[i]nadequate communication and a lack of infor-*

The same considerations hold true for a comparable but slightly different approach that emerged during the third era and that is now widely established within organizations: the involvement of representatives from different domains of the organization into a centralized coordination process. The ISO standard 17799, for example, recommends information security activities to *"be co-ordinated by representatives from different parts of the organization with relevant roles and job-functions"* (ISO / IEC 2005a, p. 10).[37] From the abstract perspective taken herein, this suggested collaboration is unquestionably aimed at decreasing efficiency losses. It reduces uncertainties and lack of information. On the other hand, content losses and alterations can be expected for the communication process, too. Consultations with and involvement of different players from throughout the organization can thus be interpreted as further, delegation-like[38] approach to minimize efficiency losses within the hierarchical coordination of security-related member behavior.

Altogether, different approaches of generalization and delegation have still been used during the third era to optimize the process of coordinating security-related member behavior. Existing methods from the second era have been retained and adapted to the specific requirements arising from the use of PCs instead of centralized mainframes. Similar to the optimizations of the second era, this allowed to reduce the number of relations having to be considered and lessened the central instance's level of uncertainty and lack of information, thereby reducing efficiency losses. But at the same time, all approaches also brought with them some secondary losses arising in the form of over-generalization or control-loss.

Besides the abstract optimization concepts of generalization and delegation already known from the second era, there is also a third approach that especially addresses the specific characteristics of the third era. As mentioned above, most of the inefficiencies of hierarchical coordination within the PC-based environments arise from the high level of uncertainty the central instance has to cope with. Besides *generalization* and *delegation*, one could thus also try to minimize the underlying information asymmetries directly and without further organizational constructs like delegation. During the third era, this direct minimization of asymmetries was increasingly realized by means of *technology*.

As outlined in section 3.4, technological progress can affect the cost structures of cooperation in different ways. Besides the size-decreasing effect of lowering communication or market costs, technological developments can also reduce the costs of *hierarchical* cooperation.[39] For the coordination of security-related member behavior during the third era, technical solutions like enterprise systems management suites or even the enterprise mechanisms included in current operating systems have exactly

mation can lead to IT security problems, wrong decisions and unnecessary working steps."

[37] A comparable suggestion is also given by the BSI (2005a, p. 18): *"IT users should be involved in the implementation planning of measures so that their ideas are also considered [...]"*

[38] The similarities notwithstanding, the concepts of delegation and representation differ in a multitude of ways. Nonetheless, examining this distinction in detail would go far beyond the scope of this work.

[39] See again Coase (1937, note 31): *"[M]ost inventions will change both the costs of organizing and the costs of using the price mechanism."*

this effect.

For instance, even the simplest network scanners allow to monitor any system being connected to the internal infrastructure and various enterprise solutions are able to detect any device being connected to a PC. Even without further mechanisms regarding the use of systems and devices being present, these monitoring capabilities allow the central instance to eliminate the uncertainties regarding the existence of systems and devices to a large extent.[40]

Another way of minimizing inefficiencies through technical means is the employment of solutions that support the coordination process more directly. Integrated security management solutions can, by means of preprocessing and reporting, effectively heighten the coordination capacity of the central instance, thereby limiting the problem of capacity limits to a certain extent. Information relevant for the coordination process can be communicated over different hierarchy-levels more accurately within integrated technical environments, leading to less content losses and alterations and thus to a reduction of control loss. And through technology-based modeling of interdependencies between different relations[41] can the effective complexity of each decision be reduced and possible reconsiderations be realized at lower cost.[42] Without going into more detail, we can thus consider technical solutions supporting the coordination process in different ways as further mechanism for reducing efficiency losses of hierarchical coordination.

Now let us, as a final step, round up the efficiency losses arising for the coordination of security-related member behavior during the third era with all the aforementioned mechanisms of optimization being present. In the idealized but hypothetical model above, we considered the number of relevant relations and thus the number of decisions having to be made as being on a high to very high level. As explicated above, this level can be lessened through different kinds of generalization. Nonetheless, the number of relations within PC-based environments will, due to the additional entities, typically still be higher than for the mainframe-based environments of the second era.[43] We will thus consider the number of decisions having to be made as *medium to high* for the optimized case.

The need for repeated reconsideration does – as mentioned above – not differ much from that of the second era and there are no optimization mechanisms aimed at changing its *medium* level.[44] Contrary to the need for reconsideration, the high to very high

[40] The widely practiced generalizations of excluding any "unknown system" from connecting to the internal network and of blocking any "unknown device" might also be interpreted as minimizing uncertainty because it leads to any system or any device having to be made known to the central instance explicitly in order to use it efficiently.

[41] A very simple method for such modeling are the attack trees introduced by Schneier (2000, pp. 318 ff). With technological support, much more sophisticated models are possible.

[42] See, for example, Schneier (2000, p. 333), highlighting that attack trees *"capture knowledge in a* reusable *form"* (emph. added).

[43] Of course, generalization could also be carried on until a similar level of relations is reached, but this would result in substantial costs of over-generalization. As we assume optimization to be applied in a deliberate manner, we will ignore this option.

[44] Nonetheless, we mentioned that supporting the coordination process by technical means might also reduce the costs for such reconsiderations. This should be kept in mind but will presumably not affect overall efficiency losses in a manner comparable to the other optimization methods as long

complexity of each decision was considered as one of the main sources of efficiency losses for the coordination of security-related member behavior during the third era. As a result to the lessened amount of relations in general and thus the lessened number of relations having to be taken into account in matters of possible interdependencies, this complexity could be lessened slightly by means of optimization. Together with the effect of technical solutions reducing the *effective* complexity of decisions, this lets us assume a *high* level of complexity for the optimized case instead of the high to very high level from the idealized basic model.

As second main source of inefficiencies besides the complexity of decisions, we identified the substantial level of uncertainty under which coordination has to be realized within the PC-based environments of the third era. Even if this uncertainty can be lowered by mechanisms like generalization over unknown entities, delegation, the involvement of representatives or through technical means, it will nonetheless always stay above the uncertainty and lack of information that prevailed during the second era. As long as PCs are not used as overdimensioned terminals, the increased flexibility will always take its toll. The level of uncertainty is thus considered as *high* for the optimized model of the third era.

And finally, we have to estimate the additional costs arising from the different approaches of optimization. The most important factor here are the costs arising from over-generalization. Even if generalization allows to lessen most other factors determining efficiency losses, it will at the same time always limit possibilities of flexible adaptation to specific requirements and thereby lessen the benefits that could basically be generated from the use of PCs instead of mainframes. Further drawbacks arise from the well-known problem of control loss and from the need to purchase and operate additional technical solutions. Summed up, these drawbacks can be considered to be substantially higher than those arising from optimization during the second era. We will thus consider them as being on a *high* level.

Taken altogether, this leads us to a *(medium to) high* level of inefficiencies having to be considered – and to be accepted by organizations – for the process of coordinating security-related member behavior during the third era of PC-based, distributed environments (see table 6.5).

as the need for repeated reconsideration stays on a medium level.

Table 6.5: Efficiency losses of hierarchical coordination – Decentralized systems

	Era 3 (decentralized systems)
Number of necessary decisions	medium – high
Need for repeated reconsideration	medium
Complexity of each decision	high
Level of uncertainty / lack of information	high
Additional drawbacks due to optimization	high
Expected level of efficiency loss	*(medium –) high*

Conclusion:
The main advantages introduced by the use of PCs instead of mainframes resulted from a significantly increased level of flexibility and user autonomy that broadened "ordinary" members' possibilities of action even further. The full potential of PCs, however, could only be tapped with PCs also being interconnected with each other to allow for the same mechanisms of information and resource sharing to be used that were already present during the second era. This interconnection together with the use of PCs instead of simple terminals led to different and much more sophisticated interdependencies being possible. Furthermore, coordination now had to be realized on the basis of even more and – in addition – partially unknown entities and activities. Even if the substantial inefficiencies resulting from these changes can again be minimized by different mechanisms (generalization, delegation, technical support), the remaining losses are still higher than those from the second era and remain on a comparably *(medium to) high* level.

6.4 Bureaucracy Costs

Besides the costs arising from efficiency losses, we also identified *bureaucracy costs* as additional factor determining the costs of coordination within a model of hierarchical cooperation in section 3.2.1. These bureaucracy costs refer to the problem of decision-makers realizing worse coordination than it would be necessary under the conditions of existing efficiency losses described above – an effect arising because of a well-intentioned but failing instrumental propensity to (over)manage and because of decisions not only being made on the basis of factual aspects but – at least also – on the basis of strategic considerations (strategic propensity to manage).

According to Williamson (1985, p. 149), the *instrumental propensity to manage* typically arises from an attitude of the respective coordinating instance to overestimate

the own capacity of handling complexity, leading to capacity limits (see above) being reached or exceeded. Instead of partially delegating the coordination to lower layers or deciding not to engage in a certain activity because the needed effort of coordination would realistically be too high, managers nonetheless try to accomplish that coordination, leading to worse outcomes of coordination than necessary.

The transfer to the field of coordinating security-related member behavior inside organizations is quite simple. Instead of partially delegating the task of coordination to lower layers, the central instance could overestimate its own capacity and adhere to a completely centralized mode of coordination even if it would be advantageous for the organization to delegate some decisions to site-specific security officers, for example. Another example would be the introduction of a new technology – equipping managers with notebooks, for instance – that would require substantial considerations with regard to the overall information security of the organization. The central instance could in this case overestimate its own coordination capacities (or, with the same effect, underestimate the necessary coordination effort) and wrongly come to the conclusion that the introduction would be manageable from the perspective of information security.

In contrast to this instrumental propensity stands the *strategic propensity to manage*. Different from the former case, the losses for the organization are accepted intentionally by the coordinating instance here. Even if, for instance, a decision maker knows that expanding his area of responsibility would lead to inefficiencies (because of his capacity limits being reached), he consciously accepts the respective costs arising for the organization in order to serve his own goals. Instead of aspiring the best overall outcome for the organization during the coordination process, decision makers do in this case coordinate in a way that (also) provides them with individual benefits. We thus see a classical principal-agent-relationship[45] between the organization as a whole and the coordinating instance with the latter having – as the agent – incentives to behave opportunistically.[46]

This strategic propensity can appear in a multitude of forms. There can be incentives for managers to enlarge their scope of decision making beyond the economically reasonable size to gain the individual benefits that result from larger responsibilities. Other incentives might encourage coordinators to gather more or less information than would be expedient[47], to make decisions on the basis of personal preferences instead of conscious consideration[48], or to make decisions that provide, at the expense of the organization as a whole, individual safeguards protecting the coordinators themselves against unforeseen events.

Such a strategic propensity to manage is well-known in the field of organization-internal information security, too: Instead of determining the most efficient outcome of security-related member behavior on the sole basis of objective risk-benefit consid-

[45] See section 3.2.2.

[46] Gurbaxani and Kemerer (1990) make the same point from the perspective of information systems management in general.

[47] See, for example, Feldman and March (1981).

[48] See Williamson (1985, p. 149): *"If [...] pecuniary incentives in firms are weaker than those in markets, then [...] preferences have greater sway."*

erations, responsible persons have an incentive to coordinate in a way that prevents them from being held responsible in the case of adverse events. It could, for example, be absolutely rational from the individual standpoint of a coordinator to decide that external devices with storage capabilities "should not" be used even if rational considerations would have come to the conclusion that the overall benefits of using them would exceed the respective overall risk for the organization. By deciding against the collective interest of the organization, the coordinator could in this case serve his individual goal of safeguarding his position.[49] Further examples of a strategic propensity to manage in the field of information security include coordination outcomes that prescribe the involvement of information security personnel for a multitude of activities (like copying files from and to the internal network), thereby heightening the importance of the security department, or, again, a central instance making less use of delegation than would be reasonable just to gather or keep a high level of impact and significance within the organization.[50]

Besides the instrumental and the strategic propensity to manage, economic theory mentions several further effects influencing the costs of coordination in a similar manner. These shall not be discussed in detail here. For our purposes, it is enough to recognize that the organization as a whole and the instance that coordinates security-related member behavior basically are in a principal-agent-relationship. With the above considerations and the knowledge about principal-agent-relations as outlined in section 3.2.2 in mind, we can thus assume that the coordinating instance will not always perform in the best interest of the organization as a whole but will rather behave in an opportunistic manner and pursue own goals, too. The outcome of the coordination process can thus be expected to be always less efficient than would be possible with the constraints given by existing asymmetries and uncertainties outlined above. This raises additional costs of coordinations on top of the efficiency losses considered in sections 6.1 to 6.3.

Even if reliable quantitative statements about these losses can hardly be made, it can be assumed on the basis of general principal-agent considerations that they basically increase with growing information asymmetries, uncertainties and complexity forming the basis for the coordination process. The more complex considerations are, the less can the outcome of the coordination process be (in)validated and the higher are the incentives for the coordinating instance to pursue own goals or, more abstractly, to behave opportunistically. The higher the uncertainty under which coordination has to be realized, the more is the coordinating instance motivated to safeguard the own position. And the less information is known, the higher are the incentives to gather disproportionate amounts of information at reasonable cost.[51] We can thus assume

[49] Schneier (2007c) refers to this phenomenon as *"CYA Security"*. A comparable class of strategic decisions which is, however, not clearly and exclusively associated with the process of coordinating member behavior was mentioned by Anderson (2001, p. 6): *"[M]anagers often buy products and services which they know to be suboptimal or even defective, but which are from big name suppliers. This is known to minimize the likelihood of getting fired when things go wrong."*

[50] See also Gurbaxani and Kemerer (1990, p. 282): *"For example, the salaries of these managers are often related to the scale of the operation, inducing them to indulge in so-called 'empire building.'"*

[51] See again Feldman and March (1981, p. 183) noting that the incentives to gather too much information are positively affected by factors like ambiguity of decisions, vague performance measure-

bureaucracy costs to have tendentially risen from the first to the third era and will consider the above-mentioned aspects as additional factors determining the costs of coordinating security-related member behavior over the different eras in our following preliminary conclusion.

Conclusion:
Bureaucracy costs arise as further costs of hierarchical coordination on top of efficiency losses. They are caused by the coordinating instance overestimating its capacity of coordination and – more important - from goal divergence between the coordinating instance and the organization as a whole. Together, these effects lead to an outcome of the coordination process that is – from the viewpoint of the organization – worse than necessary under the conditions given by efficiency losses of hierarchical coordination. The resulting bureaucracy costs increase with rising asymmetries and uncertainties and have to be considered as additional factors of the costs of coordinating security-related member behavior hierarchically.

6.5 Preliminary Conclusion

So far, we have discussed the process of *coordinating* security-related behavior of individual members of an organization in a hierarchical manner. We elaborated on different kinds of costs arising in this connection – in particular, efficiency losses and bureaucracy costs – and found these costs to have changed significantly with the repeated changes of prevailing paradigms of computer usage inside organizations. With every step – from isolated systems over mainframe-based, centralized environments to PC-based, decentralized settings – the costs arising from the coordination process increased further.

This development was primarily driven by ever-increasing possibilities of action for individual members which led to constantly increasing asymmetries between a central, coordinating instance and the actual conditions of computer usage within the organization. Even if optimization mechanisms like generalization, delegation and, especially during the third era, technological support for realizing the coordination process allowed to limit this development to a certain extent, the remaining losses nonetheless increased from era to era.

Besides these efficiency losses, bureaucracy costs have to be considered as heightening the costs of hierarchical coordination even further. Due to a tendency to overestimate own coordination capacities and because of pursuing own goals, the coordinating instance typically realizes worse outcomes than would be necessary. Like it is the case for the efficiency losses, these bureaucracy costs tend to increase with increasing asymmetries, complexity and uncertainty. They can thus also be assumed to have risen from era to era and we consider them them as having changed from a nearly non-existent to an at least medium level from the first to the third era. Altogether, we can thus as-

ments, or substantial interrelations between different uncertain decisions, for example.

sume the costs of coordination to have constantly risen from a low level during the first era of isolated systems to a high level being present in the third era of decentralized PCs. Table 6.6 sums up these findings.

Table 6.6: Costs of hierarchical coordination – Summary

	Era 1	Era 2	Era 3
Number of necessary decisions	low	medium	medium – high
Need for repeated reconsideration	low	medium	medium
Complexity of each decision	low	medium	high
Level of uncertainty / lack of information	low	low – medium	high
Additional drawbacks due to optimization	–	low – medium	high
Overall efficiency loss	*low*	*(low –) medium*	*(medium –) high*
Bureaucracy costs	–	low	medium
Overall costs of hierarchical coordination	*low*	*medium*	*high*

These conclusions shall be understood in an abstract, broad and generalized sense, illuminating the inherent influence of different computing paradigms on the process of coordinating security-related member behavior under conditions of hierarchical cooperation. Some actual, "real world" settings might not conform with or even contradict our conclusions, but in general, the continuous trend toward providing members of an organization with ever-increasing possibilities of action also leads to ever-increasing inefficiencies and misallocations characterizing the hierarchical coordination process. Even if one assumes the coordinating instance *not* to behave opportunistically but rather to be a perfect agent of the overall organization, the ever-increasing efficiency losses would still result in the same abstract principle of ever increasing costs of coordination as long as there are no optimization mechanisms in place that allow to compensate the additional losses entirely.

Notably, this does not mean that organizations were better off during earlier eras because of smaller losses of coordination. Instead, every change from one era to another allowed them to generate more value because of members being able to use computers in new ways, thereby realizing a higher level of overall productivity. But as a matter of principle, the imperfectness of the coordination process increased from era to era and organizations had to accept increasing losses as a result of the determined state of member-behavior being decreasingly aligned with the hypothetical state of "optimal" behavior. With every step forward, hierarchical organizations were less and less able

to exactly determine what members "should" and what they "should not" do.

What, then, do these abstract findings imply for the "real" world of information security management inside organizations?

First, our theoretical model developed so far suggests that any attempt to even rudimentarily reach "perfect" coordination of security-related member behavior would for many cases be inefficient within the currently prevailing paradigm of decentralized, PC-based computer usage. Whatever the result of a coordination process within a real, hierarchically structured, and sufficiently large organization is, it basically has to be considered as being substantially suboptimal. Independent from the applied mechanisms of motivation, organizations thus have to accept considerable losses arising from an "imperfect" coordination of security-related member behavior.

Furthermore, the abstract considerations might provide at least *one* explanation for the fact of not all organizations having made the step from the second to the third era and of others actually beginning to return to the model of the second era after having used decentralized, PC-based environments for a while. If an organization does not profit enough from the increased flexibility and from the members' broadened possibilities of action to counterbalance the increased losses arising from the inherent characteristics of third-era environments, then staying with or returning to the second-era model is advantageous. Interestingly, empirical observations seem to support this correlation. Those organizations which still or again prefer mainframe-like approaches over flexible, PC-based environments include, for example, banks, insurance companies, or specific areas of public administration. In all these fields, member activities are highly structured and well-defined and therefore require no or only marginal flexibility. It thus seems plausible to argue that an increased flexibility would not provide enough benefits to counterbalance the coordination losses that would arise from using PCs instead. There will of course be a multitude of further explanations but it might be interesting to examine this relation further and more formally.

And finally, our model does at this stage already suggest possible approaches for practical enhancements of information security inside organizations. As delineated above, all coordination inefficiencies have their root in information asymmetries, uncertainty, and in the capacity limits of the coordinating instance. As the various mechanisms of optimization discussed in this section have shown, these adverse conditions do not have to be taken for granted. Instead, it is possible to counteract them – at least to a certain extent – and thereby to minimize the losses of hierarchical coordination.

Especially for the third era, we mentioned technical means to be used for optimization. Network scanners allow to minimize uncertainty about systems being connected to the internal infrastructure, enterprise management solutions realize the same for devices being connected to systems, and other applications support the coordination process, heighten the effective capacity of the coordinating instance and thereby reduce efficiency losses arising from capacity limits being reached. If such solutions and other applications also counteracting the root cause of efficiency losses (and, indirectly, of bureaucracy costs, too) were developed further, this would allow for an increasingly improved coordination process and to losses being substantially alleviated. But as long as asymmetries, uncertainty and effective complexity stay where they are nowadays,

the mechanisms of motivation can be as good as one likes. They could even be "perfect". The organization nonetheless had to bear substantial losses. Future technical research in the field of information security within organizations should thus not lose sight of the coordination aspect and try to find ways for counteracting the mentioned root causes.

We will return to these and further implications of our abstract considerations later in this work. For now, let us finish considering the hierarchical coordination process and proceed to the complementary task of motivation.

Conclusion:
Because of ever-increasing information asymmetries, uncertainty and complexity, the hierarchical process of coordinating security-related member behavior gave rise to ever-increasing losses from era to era. Organizations have to arrange with these losses if they want to profit from the increased overall benefits made possible by technological advancements. Nonetheless, different approaches exist to decrease the losses arising during the coordination process and technical support might play an important role here. Future technical research in the field of organization-internal information security should therefore also concentrate on supporting the coordination process by counteracting the root causes of losses: asymmetries, uncertainty and complexity.

Chapter 7

Information Security and Hierarchical Motivation Costs

Chapter 7

Information Security and Hierarchical Motivation Costs

Management means, in the last analysis, [...] the substitution of responsibility for obedience to rank [...].

– Peter F. Drucker

Up to now, we have solely concentrated on the *coordination* process of determining the state of member behavior that would be most advantageous for the whole organization. Nonetheless, coordination alone is not sufficient to reach cooperation among different players. As outlined in chapter 3, any system for fostering cooperation also has to solve the task of *motivation*. In this chapter, we will thus examine the process of hierarchical motivation with regard to security-related member behavior.

Let us therefore assume that there is a state of security-related member behavior that was – with all the inefficiencies outlined above – determined by the coordinating instance and that shall now be realized. There are different mechanisms for doing so, but to treat the aspect of motivation as abstractly and theory-based as it was done for the task of coordination above, we first need an abstract model for the relation between the organization and the single member. The concept of *principal-agent relations* provides such an abstraction.

As outlined in section 3.2.2, principal-agent relations describe situations with one player (the agent) performing a certain task on behalf of another player (the principal) under conditions of information asymmetries and goal-divergence. Generally speaking, these conditions raise the problem of opportunism – the possibility of the agent (also) pursuing own goals instead of (merely) those of the principal – and the problem of moral hazard – the opportunity for an agent to ascribe good outcomes of his doing to high efforts and bad outcomes to adverse conditions.

The applicability of agency theory to the field of computer usage inside organizations was already discussed in brief by Gurbaxani and Kemerer (1990). With a focus on end-user computing in general, they noted the facts of the organization (top management), the IS function and ordinary members usually pursuing divergent goals and of information asymmetries existing between these players. Gurbaxani and Kemerer highlight the existence of three different kinds of principal-agent relations in this context:

- The relation between top management (P) and the functional areas (A) that should engage in productive activities and generate value,
- the relation between top management (P) and the IT-department (A) which should carry out all aspects of operating the organization's information technology, and
- the relation between functional areas (P) and the IT department (A) which should in this case provide information technology and services to the different users in accordance with their needs.

These three relations can also be identified for the more specific field of information security. In our more abstract model, the "top management" used by Gurbaxani and Kemerer corresponds to the abstract concept of the organization as a whole, the IT-department refers to the instance being responsible for information security (a CISO, for example), and the "functional units" basically equal our notion of individual "ordinary" members. The only significant difference between our notion and the model derived by Gurbaxani and Kemerer regards the relation between the instance being responsible for overall information security and the ordinary members: Instead of representing the agent, the CISO or the information security department is, for the context of realizing information security within the organization, in the role of the *principal* having to motivate a certain kind of behavior on the part of individual members. This results in the structure of principal-agent relations that is depicted in figure 7.1.

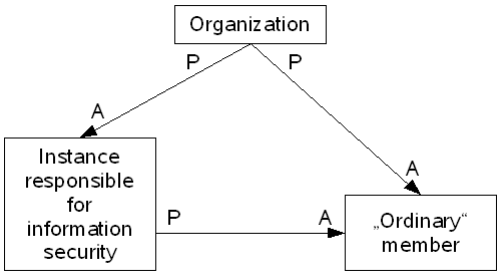


Figure 7.1: Principal-agent-relations with regard to information security (adopted from Gurbaxani and Kemerer 1990, p. 282)

For our further considerations, we will – to make the general points more clear again – solely focus on the relation between the individual members and the instance being responsible for information security.[1] It is clear that this relation meets all characteristics of a principal-agent relationship: We have a goal that the principal aspires to reach (achieve the member behavior that was determined during the coordination process), the outcome strongly depends on individual behavior of the agent, principal and

[1] To keep considerations as lucid as possible, we will also assume that the coordinating instance from above and the motivating instance are basically the same.

agent can be assumed to have divergent goals, and there are information asymmetries between principal and agent regarding the agent's current situation, his requirements arising from this situation and his actual behavior. The problem for the instance being responsible for information security (the principal) is, then, to make the individual member (the agent) actually behave in the desired manner.

To reach this goal, classical agency theory almost entirely relies on monetary incentives being directly or indirectly employed to effectively eliminate the goal divergence between principal and agent.[2] Mechanisms like performance-based payment or the deposit of bonds are usually considered here. However, human behavior can not only be influenced through economic incentives. And interestingly, neither compensation systems nor bonding schemes currently play a significant role within the field of organization-internal information security. It is thus the question how members of an organization can be motivated to behave in accordance to the state that was determined during the process of coordination.

This is where our "security triangle" comes into play. In section 2.2, we identified the three meta-measures of "architectural means", "formal rules" and "informal rules" and characterized them in terms of their strictness and their enforcement approach (ex-ante vs. ex-post). Within our model of coordination and motivation, these meta-measures primarily[3] serve the task of motivation. We will therefore address the different meta-measures from the security triangle with explicit regard to the task of hierarchical motivation. To do this on an abstract level, we need – besides the concept of principal-agent relationships – a more generalized understanding of the different meta-measures. This shall be developed in the following sections.

7.1 Meta-Measures and the Lessig Model

For the abstract consideration of different kinds of influence to human behavior from a computer science perspective, the most popular abstract model is that of the four modalities developed by Lawrence Lessig (1998a, b, 1999, 2006). In his taxonomy of constraints on human behavior, Lessig identifies *law*, *morms*, the *market* and *architecture* as different modalities (see figure 7.2). Laws prohibit or prescribe certain kinds of behavior and are enforced by the state through the threat of punishment. Norms within a social community also prohibit and prescribe behaviors but are enforced by the members of the community themselves and not by a central enforcer. Both, laws as well as norms, are basically enforced through sanctions being imposed *after* the violation of the respective rule (ex-post).

Market and architecture affect behavior in a different manner. In Lessig's model, the market-modality refers to monetary constraints limiting an individual's ability to consume a certain good or to engage in a certain activity and architecture stands

[2] See, for instance, the definition given by Stiglitz (2008), which primarily focuses on compensation: *"The principal-agent literature is concerned with how one individual, the principal [...], can design a compensation system (a contract) which motivates another individual, his agent [...], to act in the principal's interests."*

[3] Up to now, they could even be considered as *solely* being aimed at the task of motivation. But as we will see later, they have certain relevance for coordination, too.

for the *"features of the world"* (Lessig 1998b, p. 663) that restrict an individual's behavioral options. Both modalities affect behavior – as Lessig (1998b, p. 664) puts it – *"more directly"* than law and norms do. Instead of basically allowing to break a rule and threatening sanctions for this case, market and architecture constrain behavior in a way that limits an individual's effective options and can – at least to a certain extent – prevent rule-breaking behavior *in advance* (ex-ante).

Besides direct influence, all four modalities can also be considered as influencing each other. Law can, for example, prescribe a certain architecture that in turn influences behavior in a specific manner that was intended by law-makers. Market-given constraints might influence the norms being prevalent within a certain community and technological progress might lead to an architecture of man-made artifacts that influences norms as well as legal rules. Basically, any modality can not only constrain an individual's behavior but can also have influence on any other modality.[4] The whole model is illustrated in figure 7.2.[5]

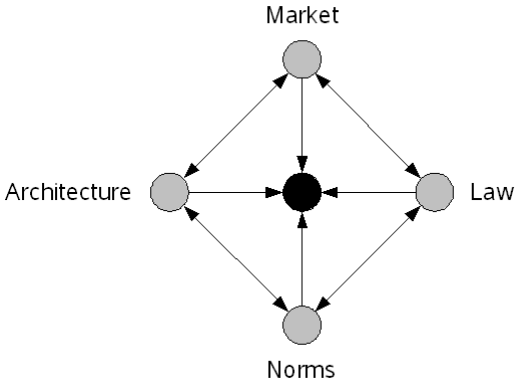


Figure 7.2: Four modalities of regulation according to Lessig (1998a, p. 664, 667; 1999, pp. 88, 93)

As some readers might already have recognized, the first three of Lessig's four constraints (architecture, law and norms) represent more general concepts of our three meta-measures identified in section 2.2.5. "Architecture" refers to architectural means being employed to realize information security, "Law" resembles the concept of organization-internal formal rules for security-related member behavior, and norms correspond to informal rules (see table 7.1). To substantiate this analogy, we shall dis-

[4] As legal scholar, Lessig is primarily interested in the role of the law. In particular, he concentrates on the possibilities of influencing behavior by means of law indirectly through legal regulation of the other modalities. Nonetheless, Lessig is also aware of the existence of further interdependencies between the different modalities. See especially Lessig (1998b, p. 667, note 22): *"[T]here is an equally important story about the market, for example, affecting other constraints, or norms or architecture as well. And with these other stories, there would be another range of arrows representing influence one way or the other."*

[5] The arrows between market and norms as well as between law and architecture are omitted for illustrative reasons alone. They nonetheless have to be considered, too.

cuss the distinct characteristics of Lessig's modalities of architecture, law and norms in more detail and with explicit regard to our three meta-measures. Even if doing this in detail would certainly require more than one book to be written for each of the constraints, some basic statements can very well be made. In particular, the modalities can, like it was done for the meta-measures in section 2.2, be distinguished by their enforcement approach and their strictness. Besides these, we will also distinguish Lessig's modalities on the basis of their *"violability"* (Surden 2007, p. 119) and their respective sources of enforcement to support subsequent discussions.

Table 7.1: Analogy between security triangle and Lessig's forces.

Meta-Measure		Corresponding Lessig-Constraint
Architectural means	$\longleftrightarrow$	Architecture
Formal rules	$\longleftrightarrow$	Law
Informal rules	$\longleftrightarrow$	Norms

Regarding the enforcement approach, ex-ante and ex-post modes of operation can basically be distinguished. In this dichotomy, law-like constraints are enforced through an *ex-post*[6] approach of sanctions being threated for the case of rule-breaking.[7] The same is true for the modality of norms which also regulate behavior through *ex-post* sanctioning after the fact.[8] Architecture-based constraints, in turn, influence behavior on an *ex-ante* basis by preventing individuals from breaking the underlying rules at all.[9] The same characteristics were also identified for the corresponding meta-measures in section 2.2: Architectural means for information security operate ex-ante while formal as well as informal rules basically follow an approach of ex-post enforcement.

Closely connected with the approach of enforcement is the general possibility of breaking the rules. Cheng (2006, p. 671) refers to this as the *"ability to break the law"* while Surden (2007, p. 119) uses the more vivid concept of the *"violability"* of a constraint – the ability of the constrained individual to disobey the rule that is to be enforced. Both authors characterize architecture-based constraints as being of a low violability in contrast to legal constraints. Lessig (1999, p. 236) makes the same point of architectural constraints not being as ignorable as laws. Even if it is generally possible to overcome the constraint, violations do in this case require much more effort than the simple ignorance of a legal rule. The violability of architectural

[6] Note that the underlying principle of ex-post enforcement is deterrence. The threat of ex-post sanctions basically serves the purpose of discouraging individuals from breaking the rules. This deterrence is in fact aimed at functioning *before* rule-breaking behavior takes place. Nonetheless, enforcement still takes place *ex-post*. See also Lessig's (1998b, pp. 667 ff) discussion on objective and subjective constraints.

[7] See, for instance, Lessig himself (1998b, p. 662), Reidenberg (1998, p. 568), Cheng (2006, p. 659) or Surden (2007, p. 121).

[8] See again Lessig (1998b, p. 664). See also Ellickson (1999, p. 4) and further sources included therein.

[9] See Reidenberg (1998, pp. 568 f), Cheng (2006, p. 662) or Surden (2007, p. 119).

constraints can thus be assumed as being low but not as non-existent. Norms can again be considered as being similar to legal rules here. Even if there is a threat of being sanctioned ex-post, it is possible for an individual to behave in non-conformance with a certain social norm. Norms are thus of high violability, too.

The transfer of these characteristics to our meta-measures is fairly trivial: Architectural means like access limitations can only be overcome with substantive efforts and have thus to be considered as being of low violability. Formal as well as informal rules can, by contrast, easily be ignored or violated and are thus of a high violability.

The property of strictness was defined by the accurateness and explicitness and by the *"possibilities for instant reaction to exceptional and unforeseen situations"* in section 2.2. A meta-measure that accurately and explicitly distinguishes between allowed and disallowed behavior and that provides few opportunities for situational flexibility was considered as being of a high strictness. Obviously, this property can also be assigned to the different constraints from Lessig's model.

Architectural constraints, for example, typically tend to regulate behavior in a *"dichotomous and binary"* (Surden 2007, p. 119) or at least – compared with law-like constraints – *"less discretionary"* (Cheng 2006, p. 665) manner. They are typically not subject to trade-offs or weightings, not to speak of situation- or context-based subjective interpretation or of unanticipated derogations. Even if the underlying rules can be considerably complex, they are basically of objective and completely defined nature and constrain individual behavior in a highly determined manner. Like the corresponding meta-measure of architectural means, Lessig's "architecture" can thus be assumed as being highly strict.

Law-like constraints are, by contrast, always subject to discretion to a certain extent. Even if laws are formulated explicitly, they are usually more vague than architecture. Due to the ex-post approach of regulation, the enforcers of possible sanctions are furthermore able to take unforeseen aspects into account and to interpret the underlying rules subjectively, leading to less determined outcomes and allowing for well-justified deviance. Laws are thus less strict than architecture. Even more than for law-like constraints, this applies to norms. These are not formulated in an explicit manner but rather informal and in many cases unarticulated, tacit and vague.[10] Again, this is consistent with the strictness-classification of the respective meta-measures developed in section 2.2.

Finally, we can also distinguish the different constraints by their source of enforcement. While architectural means can be considered as self-enforcing[11], law-like constraints require some specific enforcement instance being responsible for the detection of infringements and for realizing the subsequent sanctions. Norms, in turn, are enforced in a highly decentralized manner[12] by the members of the respective community

[10] See, for example, Ellickson (1991, p. 130), mentioning practitioners of a norm being unable to articulate the content of that norm. Of course, norms can also be well-defined, might clearly distinguish between conformance and non-conformance and can even imply a well-defined sanction to be imposed upon norm-breakers. Nonetheless, this determinedness is, from a general point of view, less distinct for norms than for law-like or for architecture-based constraints.

[11] See, for example, Reidenberg (1998, p. 568)

[12] See, for example, Ellickson (1991, p. 130).

themselves.[13] The same sources of enforcement are also present for the corresponding meta-measures. We can thus conclude that each of our meta-measures has a more generalized representation within the Lessig-model of forces influencing human behavior. Furthermore, we can add the properties of "violability" and "source of enforcement" to our model of meta-measures, resulting in the classification outlined in table 7.2.

Table 7.2: Refined classification of meta-measures.

Meta-Measure	Enforcement Approach	Strictness	Violability	Source of Enforcement
Architectural means	ex-ante	high	low	self-enforcing
Formal rules	ex-post	medium – high	high	dedicated instance
Informal rules	ex-post	low – medium	high	community-driven

Besides already representing an interesting fact on its own, this analogy also allows us to derive well-founded arguments regarding the costs arising from the different meta-measures being used to motivate members to behave in accordance with the outcome of the previous coordination process. To discuss the *motivation costs* of hierarchical cooperation in the field of organization-internal information security on the basis of principal-agent theory *and* the Lessig-model, we just have to make one more step of abstraction: As already mentioned above, principal-agent theory almost solely relies on monetary considerations. To make our model of meta-measures compatible with this approach, we will thus translate the constraining effects of the different modalities and the according meta-measures to a unitary dimension of "costs" being imposed on the influenced behavior.

Doing so might seem deceptive at first sight, but in fact, the abstract economic concepts of costs and benefits reach far beyond the scope of traditional economic activities. Gary Becker, for example, extended the economic approach to such "non-economic" subjects as discrimination, crime or family planning.[14] The underlying idea of this approach is that economic principles can be applied to any kind of human behavior, given some assumptions like the imputability of monetary values to rather emotional concepts like pleasure or pain.[15]

[13] For other kinds of norm enforcement – including emotional or self-enforcement – see, for instance, Posner (1997). We will nonetheless concentrate on the primary approach of sanctions being imposed on the norm-breaker by a community.

[14] In 1992, Becker received the *"Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel"*, which is commonly named the Nobel Prize in Economics – *"for having extended the domain of microeconomic analysis to a wide range of human behaviour and interaction, including non-market behaviour."*

[15] See especially Becker (1978, p. 8): *"Indeed, I have come to the position that the economic approach*

The application of this approach to our model is most obvious for the "law" constraint or the "formal rules", respectively. Becker (1978, pp. 39 ff), for example, presents an economic explanation for crimes being committed under threat of severe punishments. In this example, the "costs" of committing crime are basically assumed to be a function of the probability of being caught and convinced on the one and the severity of punishment on the other hand.[16] The result of this function represents the individually perceived risk of the would-be rule-breaker.[17] The legal constraint does, then, deter the individual from breaking a rule only if the individually perceived risk exceeds the estimated individual benefit of rule-breaking behavior.

For norms or "informal rules", the individual costs of disobedience can be derived in a comparable manner. Different from the "law"-constraint, norms are enforced in a decentralized way by the relevant community and it is thus this community or, more precisely, the members of the community that have to detect rule-breaking behavior and to realize punishments. Due to this different source of enforcement, methods of punishments differ strongly from those being employed by a dedicated enforcement instance in the former case.[18] These sanctions can nonetheless also be understood in terms of "costs" having to be borne by a rule-breaker in the case of being detected and penalized. And again, these costs and the probability of being sanctioned together determine the individually perceived risk that would arise from rule-breaking and that has to exceed the respective individual benefits in order to make the constraint an effective one.

For both, law-like formal as well as norm-like informal rules, the individually perceived risk – and consequently, the impact of the constraint – thus mainly depends on the estimated probability of being discovered and the severity of threatened sanctions. This is different from the third constraint of architecture-based influence. Due to the approach of self-enforcement being realized ex-ante and the low violability, the individual costs of disobedience arise in a different manner here. Instead of being the result of a function of probability and severity of sanctions, the costs of rule-breaking do in this case represent the effort having to be made to overcome the constraint. If this necessary effort is higher than the individual value derivable from a certain activity, the overall payoff of that activity is negative and again, the individual would be restrained from the respective behavior.

---

is a comprehensive one that is applicable to all human behavior, be it behavior involving money prices or imputed shadow prices [...]." See also Becker (1993).

[16] As mentioned above, the severity of punishment can be valued in monetary terms even if the punishment itself is not of monetary nature. The mentioned function is not a simple mathematical product of probability and severity, as in this case, one would be able to reach considerable "costs" to the criminal with a minimal probability of being caught and with huge sanctions for the (unlikely) case of being punished. In fact, Becker's calculations are much more sophisticated, also including aspects like risk-aversion or -preference, and result in the normative statement that probability and severity should be set to *"those regions where offenders are, on balance, risk preferrers"* because *"crime does not pay"* in these regions (Becker 1978, p. 53).

[17] Generally speaking, sanctions could also be positive and reward rule-conformance instead. Even if there are notable differences between negative and positive sanctions, we will only consider punishments to simplify matters again.

[18] Ellickson (1999, p. 8), for example, explicitly mentions gossip, ostracism and adjustment of exchange relationships as possible sanctions.

The most important point here is that architectural means allow the "costs" of rule-breaking to be certainly high. Even if only few architecture-based constraints are absolute and would thereby raise "infinite costs" having to be borne to engage in rule-breaking behavior, it is nonetheless possible to realize considerable burdens that are perceived as being "nearly infinite" by most of individuals. The costs being induced by architecture-based constraints do not depend on detection and sanction but merely on the nature of the instrument itself.

In most cases, however, all forces influence the individual simultaneously and the actual behavior depends on the aggregated effect of all constraints being at work.[19] If the resulting aggregated "costs" of a certain behavior exceed the individual benefit, then the overall payoff for the individual to act in this manner would be negative and consequently, the behavior would not take place.

All constraints or meta-measures can thus be used separately as well as in combination with each other to change the economic conditions faced by the individual. They thereby allow to exert influence on behavior and can be used for the purpose of motivation. With this final abstraction of understanding constraints as "costs" being associated with the possibility of rule-breaking, we can now proceed with our considerations of hierarchical motivation costs.

7.2 Agency Costs and Meta-Measures

As mentioned above, we can understand the relation between the instance being responsible for information security and the individual, "ordinary" member as principal-agent-relationship. The problem having to be solved within this setting is that of goal divergence between principal and agent resulting in agent behavior that does not conform to the interest of the principal and that – indirectly – also conflicts with the collective interest of the organization as a whole. Different approaches for overcoming this problem exist, but all of them also cause certain costs. These costs shall now be discussed in more detail and with explicit reference to the task of motivating "secure behavior" among the members of an organization.

Classical agency-theory distinguishes three kinds of costs: Monitoring costs, bonding costs, and the residual loss (Jensen and Meckling 1976). Within this taxonomy, monitoring costs do not only refer to the costs having to be spent by the principal for observation purposes but also to the costs of exerting influence on the agent.[20] This broad concept of monitoring costs represents – together with the remaining losses – the costs usually arising from different mechanisms being used in order to influence the

[19] Note that different constraints might also contradict each other. See, for instance, Lessig (1999, p. 87): *"Changes in any one will affect the regulation of the whole. Some constraints will support others; some may undermine others. A complete view, however, should consider them together."*

[20] See Jensen and Meckling (1976, note 9): *"[T]he term monitoring includes more than just measuring or observing the behavior of the agent. It includes efforts on the part of the principal to 'control' the behavior of the agent through budget restrictions, compensation policies, operating rules, etc."* The costs of *"devising and enforcing specific behavioral rules or policies"* are also mentioned explicitly by Jensen and Meckling.

agent's security-related behavior.[21] With our meta-measures and their more detailed characterization developed above, we have three different classes of such mechanisms all of which feature different cost structures. These shall therefore be analyzed in the following. Like in the case coordination costs, reflections will again be made on a generalized, abstract level to make the central aspects more clear. Remember that we still assume that a desired state of security-related member behavior was already determined and that this state should now be realized.

7.2.1 Costs of Architectural Means

Once an aspired state of security-related member behavior is determined and completely specified, architectural means seem to be predestined for enforcing this state throughout the organization. Different from the mechanisms typically considered in principal-agent theory, architectural means can be used to impose virtually infinite costs on rule-breaking behavior and thus to achieve rule obedience even for cases where individual members could derive substantial benefits from rule-breaking. Overcoming an access control mechanism that is used to enforce a previously determined state of member behavior would – depending on the implementation of the mechanism and on the member's skills, of course – require the member to make extreme efforts. These necessary efforts would then render rule-breaking disproportionately expensive for the majority of cases and consequently, the member would abstain from this effectively inefficient option.

The same principle basically applies to all established architectural means for realizing secure member behavior. Physical walls and locks lead to substantive efforts being necessary to access the respective systems, firewalls make unwanted access of external online resources significantly more expensive[22] and DRM-systems place the same burdens on unwanted access to the protected information. Technical solutions restricting the use of USB-devices impose significant costs on the usage of unapproved devices. And finally, a technical solution preventing a mobile user from connecting to public WLAN-hotspots also places a nearly infinite burden onto the option of doing so, thereby reliably enforcing member behavior that conforms to the previously determined "optimal" state. In all cases, the costs of rule-breaking that are imposed on the would-be rule-breaker are extremely high and thereby make disobedience economically inefficient for the member even if he could derive substantive benefits from rule-breaking.

What, then, are the costs of realizing motivation by means of such architectural constraints? Basically, these are the costs that were mentioned as "direct costs" of information security in section 4.2: The costs for the technical artifacts themselves and the costs arising from their operation and maintenance.[23] These two factors, in turn,

[21] Even if one could think of a wide variety of bonding agreements that could be used within this field, such mechanisms do not play a significant role for the current practice of information security inside organizations. We will thus not consider them further. Bonding agreements with regard to security-related behavior could nonetheless represent an interesting field for future research.

[22] The same is, of course, true for unwanted external access to internal resources. Architectural means are thus also able to influence the behavior of non-members.

[23] "Indirect costs", the costs arising from the impairment of daily work, are not part of the motivation

differ slightly in their nature. The expenditures for acquiring, constructing and / or installing the physical or technical means have to be borne only once for the whole use of the respective motivational instrument. They can thus be seen as *fixed costs*. The costs of operation and maintenance, on the other hand, arise continuously. If, for instance, the determined state of "optimal" member behavior changes because of repeated reconsiderations being necessary (see sections 6.2 and 6.3), repeated motivation costs arise because of the need for reconfiguring the respective technical solutions (changing access rights, for example). These repeated costs can be seen as the *marginal costs* of architectural means. The ultimate enforcement, however, is realized at zero or negligible costs because of the self-enforcing nature of architectural means.[24]

If, for example, a determined state of access and non-access to online resources is realized by means of a firewall, this firewall has to be acquired and installed once and it has to be reconfigured whenever the desired state of member behavior changes because of reconsiderations in the coordination process. The ultimate enforcement is in turn realized at minimal costs by the technical artifact. Classical file access rights mechanisms share the same cost structure. As such mechanisms are nowadays typically included in the operating system, the acquisition costs can be omitted here and the (broadly understood) monitoring costs solely arise from the need for (re-)configuration of access control lists. And finally, technical mechanisms for enforcing a certain state of WLAN-usage also raise fixed costs of acquisition and initial setup and marginal costs of reconfiguration – if a certain member should not use WLAN anymore, for example – while the ultimate enforcement for any single case is again realized at zero or insignificant costs. The monitoring costs of architectural means thus consist of some initial, fixed costs and some marginal costs having to be borne for every case of reconsideration or rather for every case of the outcome of the coordination process being changed.

Besides the broadly defined monitoring costs, agency theory suggests the existence of remaining losses. For our subject, these would be the costs arising from member behavior that differs from the "optimal" state that was determined during the coordination process. With architectural means being used for motivation purposes, these losses only have to be expected on a minor level as the low violability almost completely prevents non-conforming behavior. In fact, it could also be argued that technical means change the characteristics of the relation between principal and agent in a more fundamental manner: While classical agency theory assumes the principal to delegate a certain amount of decision making to the agent[25], architectural means and their low violability change the effective amount of decision rights held by the agent to nearly zero. In this case, it could be objected that we do not even have a

costs but do rather – if representing cooperation costs at all – belong to the *coordination* costs. In an unrealistic case of perfect coordination, an organization would be able to eliminate all productivity infringements that do not provide a larger direct or indirect benefit. The remaining, still avoided activities must then represent an overall negative payoff if perfect coordination is assumed and thus *should* indeed be avoided from a collective point of view. Any other impairment not being justified by larger risk reductions must thus be subject to imperfect *coordination*.

[24] See, for example, Cheng (2006, p. 670), noting that architectural means are *"efficient and relatively inexpensive because they do not require enforcement on a case-by-case basis."*

[25] See, for example, Jensen and Meckling (1976).

principal-agent relation anymore.

This possible plea notwithstanding, architectural means can – from an abstract point of view – be considered as a relatively cheap way of motivation within principal-agent relationships. No or only marginal residual losses have to be borne, self-enforcement makes explicit observation efforts and the execution of sanctions unnecessary and the only costs that repeatedly arise are those emerging from changed results of the coordination process. Besides these, the only significant cost factor are the expenditures for the initial acquisition and setup of the respective artifacts. Especially with the determined state of member-behavior having to be enforced in a multitude of cases – thousands of file-accesses a day, for example – architectural means thus entail low costs for each single act of enforcement.

As we will see in chapter 7.3, the use of architectural means can nonetheless result in substantive costs that can render the use of alternative constraints more efficient.

Table 7.3: Motivation costs of architectural means.

Meta-Measure	Fixed Costs	Marginal Costs	Enforcement Costs (single case)	Residual Loss
Architectural means	high	medium	none / negligible	none / negligible

Conclusion:
Architectural means are one instrument for motivating members to behave in accordance to the state of behavior previously determined during the coordination process. They are self-enforcing and therefore raise no or only negligible enforcement costs for the single case and avert residual losses. The initial fixed costs for erecting the constraint can, however, be substantial and the only source of repeated costs are reconsiderations that change the coordination outcome and thereby necessitate reconfigurations of the respective architectural means.

7.2.2 Costs of Formal Rules

Different from architectural means, motivation by law-like constraints represents the classical principal-agent relationship more consistently. Due to the high violability of this constraint, the ultimate decision about complying with the state determined during the coordination process or not lies with the agent. Law-like constraints nonetheless allow to influence this decision by changing the effective payoff being assigned to each behavioral option. To realize this kind of influence, the principal has to engage in the classical activities of discovering rule-breaking behavior and of exerting sanctions

in the case of such behavior being recognized.[26] Both of these activities raise costs that belong to the broad concept of monitoring costs mentioned above.

These monitoring costs basically arise for any single act of motivating behavior that conforms to the coordination outcome. To reach a certain amount of individually perceived risk at the side of would-be rule-breakers, any rule-breaking behavior had to be discovered with a certain probability and sanctions of a certain severity had to be realized for the case of such behavior being detected. This necessity renders the enforcement of law-like rules more expensive than architecture-based influence with regard to the single case.

If, for example, a certain state of member access to internal resources should be enforced by law-like formal rules instead of technical access control, there had to be a certain probability of unwanted access being detected and a certain severity of the respective sanction in order to reach a considerable individually perceived risk and to eventually render the activity unprofitable for the individual member. The same is true for the motivation of a determined state of access to external online resources. With a low chance of being detected or with negligible sanctions being threatened, formal rules would only hardly discourage the individual from breaking the respective rule. The transfer to the use of unapproved USB-devices is trivial and even with regard to the use of public WLAN-hotspots, a formal rule to enforce a certain state of behavior would only be effective with the likeliness of discovery and the threatened sanction being high enough. Of course, realizing a certain detection probability and exerting severe sanction raises a certain amount of costs having to be borne by the motivating instance. Motivation by formal rules thus entails costs on the case-by-case basis that are not or only insignificantly present with architecture-based motivation.

For all the cases mentioned above, it can furthermore be assumed that there will be a certain amount of members not behaving in the intended manner. Like it is the case with most "usual" legal rules, it would be economically inefficient to strive for an outcome of absolute obedience in the field of security-related member behavior because this would require extensive investments in the discovery of rule-breaking. Instead, it is in most cases more efficient to realize a lower level of rule-conforming behavior and accept a certain amount of losses arising from violations and the resulting insecurity. In classical agency theory, these are the costs referred to as *residual loss* and their existence perfectly accords with the well-known principle of "perfect" security basically being economically inefficient. Different from the case of architecture-based influence of security-related member behavior, law-like formal rules thus do not only raise the broadly understood monitoring costs but also lead to a certain amount of residual losses having to be accepted.

Law-like constraints are, however, significantly cheaper than architectural means in their initial creation. Given a specified outcome of the coordination process, law-like rules do not raise costs of acquisition and require less efforts of conscious design and implementation but can rather be formulated at comparably low costs.[27] The

[26] See, for example, Becker (1978, p. 76). For an early explicit application of the underlying deterrence theory to security-related behavior, see also Straub (1990).

[27] See again Cheng (2006, p. 670).

costs arising from their adjustment to changed coordination outcomes are, in turn, comparable to those of architectural means. There is – in matters of costs – no essential difference between changing access rights within a technical solution and changing an access prohibition in a law-like framework of rules.

The fixed costs of law-like formal rules regarding security-related member behavior are thus low while those of architecture-based regulation are high, the marginal costs of adaptation to changed coordination outcomes are medium for both cases, and the marginal costs of enforcing a certain behavior for one single case are low for architecture-based influence and comparably high for law-like rules. This leads us to a cost structure of formal rules being used to motivate security-related member behavior as summarized in table 7.4.

Table 7.4: Motivation costs of law-like formal rules.

Meta-Measure	Fixed Costs	Marginal Costs	Enforcement Costs (single case)	Residual Loss
Formal Rules	low	medium	high	medium

7.2.2.1 Optimization

Different from architectural means, law-like constraints offer a wide variety of optimizations that could also be applied to our subject of security-related member behavior. As mentioned above, the costs of law-like influence primarily arise from the need for detecting and sanctioning rule-breaking behavior.[28] We will thus consider two approaches for minimizing these broadly understood monitoring costs in brief: the concept of internalization and the possibility of supporting the enforcement process by technical means.

Of these two, the latter refers to technical (or architectural) solutions that do not directly affect individual behavior by constraining behavioral options but that do rather heighten the effective detection capabilities of the enforcing instance or, respectively, that lower the costs of observation. Examples for such solutions are logging mechanisms that record any file access of a given member or that notify the enforcing instance in case of any behavior that matches a certain predefined set of activities. Other solutions might support the detection process less explicitly by increasing the probability of rule-breaking behavior being recognized.[29] From a generalized perspective, such an influence of technical artifacts to the principal-agent relation is nothing new: Tech-

[28] The residual loss is an indirect result of these monitoring costs.

[29] Cheng (2006, p. 664) gives the example of steel mailboxes heightening the probability of would-be mail thieves to be detected and refers to this effect as *"structural laws"* of *"Type I"* while the prevention of rule-breaking behavior through architectural means are referred to as *"Type II structural laws"*.

nology allows the principal to oversee more aspects of agent behavior at lower costs and thereby minimizes the information asymmetries that shape the principal-agent relation. Applied to the field of security-related member behavior, the effects are basically the same. There also is a direct effect of lowering the costs of discovery and we can also expect an indirect effect of minimizing residual losses because of a higher probability of detection being efficiently realizable. Technology can thus not only be used as an own modality of motivation but can also serve as optimization mechanism for the meta-measure of formal rules.[30]

The second approach for minimizing the costs of law-like influence is the *internalization* of the respective formal rules by the agent. Internalization can be understood as the process of converting the external influence constituted by observation and punishment to an internal "motivation" driven by mechanisms like guilt or pride.[31] Instead of conforming to a formal rule because of the possible sanctions, the individual does in this case, simply speaking, obey the rule because "it is right" to do so. The enforcement thus needs not to be realized by some external, dedicated instance anymore but is rather done by the constrained individual herself.

For our context of security-related member behavior, such an internalization of formal, law-like rules would be present if members ascribed a personal value to rule obedience and abstained from rule-breaking behavior because they would otherwise feel guilty, for example. In this case, less investments had to be made for detection and punishment and organizations would therefore undoubtedly appreciate such an internalization. Realizing internalization does, however, at least require substantial efforts and a certain amount of time[32] both of which effectively represent additional costs. To what extent internalization of security-related formal rules qualifies as mechanism of cost optimization under such conditions has to remain unclear and might be subject to further research. Organizations might find efficient ways to reach a significantly higher level of internalization in the future and in this case, the optimization effect could in fact be significant. But for now, the impact of internalization is uncertain for the meta-measure of formal rules.

[30] The main difference between technological solutions of this kind and those discussed as "architectural means" above is that the "violability" basically remains here. Even with a detection probability of 100 percent, the individual is still able to violate the rule.

[31] Cooter, for example, uses the concept of an *"intrinsic value of obeying a law"* (2006, p. 1282) or, respectively, of guilt being attached to *"forbidden actions, thus raising their psychological costs"* (1997, p. 956). Hodgson (2006, p. 18) refers to the same mechanism from a slightly different perspective as *"habituation"*. For internalization in general, see also Lessig (1998b, p. 678).

[32] See, for example, Siponen (2000, p. 35): *"[T]here seems to be no reason for assuming that internalization of security guidelines can be easily achieved straight away [...]. Taking this into account, user acceptance and internalization must be considered gradual processes and long-term goals."* Shostack and Stewart (2008, pp. 128 ff) are even less optimistic and state – consistently with the behavioral approach pursued herein – that *"people are surprisingly resistant to indoctrination [...] where there is a* payoff for breaking the rules*"* (emph. added).

Conclusion:
Formal, law-like rules can also be used for motivating members to behave in the desired manner. Different from architectural means, formal rules can be created at low fixed costs but raise substantial marginal costs of enforcement because they require observation and punishment. Furthermore, formal rules can basically be violated and give rise to residual losses. Optimization can at least be realized through technical support of observation and through internalization of the rules by the members themselves.

7.2.3 Costs of Informal Rules

As outlined in section 2.2.3, informal rules with regard to security-related member behavior refer to those rules that are present inside an organization and influence member behavior but that are nonetheless not explicitly formulated or represented in another directly observable form. The example of "dubious" external web-sites being accessed from the internal network was already given in section 2.2.3: Even when there exists no technical representation and no formal rule about which web-sites are accepted to access and which ones not, there are still social forces that influence the individiual member in his decision whether to perform such an access or not. These social forces refer to our informal rules.

Generally speaking, such norms-like informal rules are, like law-like formal rules, enforced through ex-post sanctions. They thus require a certain probability of rule-breaking behavior to be detected and a certain severity of threatened sanctions to induce sufficient costs to the would-be rule-breaker to represent an effective deterrent. But different from law-like formal rules, informal rules are not enforced by a dedicated enforcement instance but rather by the members of the respective community themselves:[33] If one member violates an informal rule and another member observes this behavior, the latter imposes a certain sanction on the rule-breaker. Different from the concept of explicit enforcement being present for formal rules, informal rules are thus enforced through spontaneous and distributed mechanisms of *"informal control"* (Ellickson 1991, p. 131).

Consider, for example, informal rules that influence the access to internal information resources. Assume, therefore, that there is an informal rule which condemns any access to personal files of other members. Even if there exists no formal rule that prohibits such accesses, any member found accessing other members' personal files can then be punished by the "community" by means of gossip, ostracism, social exclusion or by refusals to work with the rule-breaker in future projects. If the individually perceived risk arising from these possible sanctions and from the probability of being discovered exceeds the individual benefit, this will deter the respective member from actually performing such accesses. If there existed an informal rule that disparages the use of other USB-devices than keyboards and mice or the use of public WLAN-hotspots – whatever the source of this rule might be – any member found us-

[33] Informal rules might nonetheless also guide the weighting that is performed during the enforcement of formal rules. Again, the different modalities would in this case influence each other. We will nonetheless concentrate on the *direct* effect of norms-like informal rules here.

ing external storage media or accessing WLAN-hotspots could also be sanctioned by similar means and individuals would again be discouraged from breaking the informal rule.

In all these cases, the monitoring costs of discovery and sanction need not be borne by some central instance but are rather assumed to be shouldered voluntarily by the members themselves.[34] Like architectural means, informal rules do thus raise no (broadly understood) monitoring costs that had to be borne by the central instance for any singular case of enforcing a specified behavior. From this perspective, informal rules might appear as an inexpensive mechanism for motivating members to behave in accordance to the outcome of the coordination process.

The most significant problem of using norms-like informal rules for the purpose of conscious influence regards the creation and change of these informal rules and their adaptation throughout the respective group of individuals. Instead of being promulgated by some dedicated instance like the state or, for our case, the instance being responsible for information security, norms basically *emerge* in a bottom-up direction from repeated interactions among individuals and represent a consensus of these individuals about what behavior is desired and what not.[35] Norms are thus not the result of conscious planning but rather the outcome of a complex social process. Due to this nature, the use of norms-like informal rules to influence behavior within a hierarchical principal-agent relationship is less simple than it is the case for the basically hierarchical mechanisms of architectural means or formal rules.

This does, however, not imply that there is *no* possibility for influencing the emergence of norms and their actual content in a hierarchical, top-down manner. Even if norms basically emerge bottom-up, their development can nevertheless be affected by hierarchical means that follow the top-down approach. Most of these mechanisms function through influence of change-agents – members of the social group that face an individual benefit from the norm being changed and that are by these benefits motivated to engage in norm-changing.[36] In particular, governments can change the payoff of would-be change agents – by providing subsidies or classical sanctions – and thereby provide an incentive to engage in norm changing or it can provide detailed information about the respective context in order to *"influence opinion leaders"* and

[34] See, for example, Posner and Rasmusen (1999, p. 369): *"A norm is a social rule that does not depend on government for [...] enforcement."* See also Ellickson (1999, p. 5), regarding norms as rules *"that third parties other than state agents diffusely enforce by means of social sanctions."* The motivation for members of a community to engage in enforcement and to take the respective costs is a vital subject to current research and shall not be discussed in detail here. For an introductory overview of this subject, see, for example, the various works of Ernst Fehr and Simon Gächter and, of these, especially Fehr and Gächter (2000). Generally speaking, individuals seem to be willing to punish non-cooperative behavior even if this raises certain costs and provides no individual benefit to them. Other incentives to engage in costly punishments might arise from secondary social norms, too. See also Posner (1997, p. 366), arguing that *"the costs to the enforcer will actually be negative"* in many cases.

[35] Ellickson (1991, p. 168), for example, refers to norms as *"rules created by* nonhierarchical *social forces"* (emph. added). See also Lessig (1998a, p. 21), noting that *"[n]orms can't be imposed externally"*, and recall the concept of *"tacit consent"* already used in section 2.2.3.

[36] For the role of change agents, see, for example, Ellickson (1999, pp. 15 ff).

thereby also spur the diffusion of a new or changed norm (Ellickson 1999, p. 50).[37]

To exert this *indirect* influence on the bottom-up emergence or change of norms, governments can then use their established, top-down mechanisms of influence. But different from the direct influence on individual behavior through legal rules, the initial costs of developing norms and the marginal costs of changing them are at least higher than they are for the direct case. The provision of benefits to would-be change agents or the supply of substantial information on the respective context to opinion leaders, for example, require efforts that go beyond those of a simple (re-)formulation of a legal rule. Even if the (broadly understood) monitoring costs of norms are low for any single case of enforcement, the initial costs as well as the marginal costs of changing the rule are thus at least higher than those of law-like influence.

These rather abstract considerations on the development and change of norms in general can again be mapped onto our more specific subject of organization-internal informal rules with regard to information security. While the existence of an informal rule not to access personal files of other members might possibly be assumed without explicit stimulation, this will hardly be the case for the use of unknown USB devices. To influence the use of such devices by means of an informal rule, the centralized instance thus had to foster the development of the respective rule. Due to the characteristics of social norms described above, this would require the existence of change agents that face an individual interest in the existing informal rules being changed or a new informal rule being developed, respectively.

This could, analogously to the abstract mechanisms mentioned above, be realized by providing possible change agents with an individual interest in altered informal rules. Mechanisms providing such incentives can, however, hardly be imagined for the case of using unknown USB-devices. The other option is to provide opinion-leaders with information that convinces them of the advantages of rule-change or -development and motivates them to promote the respective rule-change throughout the organization. This might, in turn, be the abstract calculus behind the conduct of awareness trainings: Members are provided with (more or less) detailed technical information about the risks arising from certain individual behaviors for the overall organization and it is hoped that there is an opinion leader being convinced of the reasonableness of existing informal rules being changed. Again, the application to the use of WLAN hotspots is trivial.

As we can see, norms-like informal rules and their use for the purpose of hierarchical motivation are much more sophisticated to analyze than architectural means or law-like formal rules. This is also the case for our subject of hierarchical motivation with regard to security-related member behavior.[38] Nonetheless, some basic properties of

[37] Ellickson (1999, pp. 50 f) also mentions the possibilities of changing the boundaries of the social community and of hampering (or supporting) the inner structures of social subgroups that engage in the development of unwanted (or desired) norms. For a government's possibilities of influencing norms, see also Posner and Rasmusen (1999, pp. 380 f), who, in particular, also mention governmental support for the various ways of informal sanctions.

[38] It is an interesting fact on its own that the importance of norms-like informal rules is widely emphasized by a multitude of authors – including scientific ones – through the use of terms like "security culture" but that at the same time the theoretical, scientific foundations are hardly ever discussed. Positive exceptions include, for instance, Schlienger and Teufel (2002), Kuusisto,

norms-like informal rules can very well be identified. We can characterize them by a low level of monitoring costs having to be borne by the central instance for any single case of rule enforcement and by at least medium monitoring costs arising for the creation and for any change of the rule being necessary because of altered outcomes of the coordination process.

Finally, norms-like informal rules are of a high violability and individual members can thus behave in an undesired manner. Furthermore, and different from the formal rules considered above, it takes significant amounts of time until a new or changed informal rule is really effective. While a technical solution regulating the use of USB-devices allows immediate changes of member behavior after the outcome of the coordination process changed and while formal rules can also be enacted and enforced promptly, the development or change of informal rules is a comparably slow process.[39] There is thus always a certain delay between a change in the outcome of the coordination process and the actual effectiveness of the changed informal rule. During this period, members face – in addition to the residual losses also being present for formal rules – a motivational situation that does not conform with the actual intent of the centralized instance, leading to additional residual losses having to be borne by the organization.[40] Altogether, this results in a cost structure of hierarchical motivation by means of informal rules as depicted in table 7.5.

Table 7.5: Motivation costs of norms-like informal rules.

Meta-Measure	Fixed Costs	Marginal Costs	Enforcement Costs (single case)	Residual Loss
Informal Rules	medium	medium – high	low	high

7.2.3.1 Optimization

The mechanisms of optimization for norms-like informal rules are basically the same as for law-like formal rules. First, technical (or architectural) facts can change the general conditions under which motivation takes place. They can heighten the likelihood of rule-breaking behavior of one individual being recognized by the other individuals of the group and thereby increase the efficiency of the enforcement process. If, for instance, any access of one member to the personal files of another one were automatically reported to the latter, this would – given the existence of an informal rule not

Nyberg, and Virtanen (2004), or Glaser (2009). Again, this opens a wide field of opportunities for future research that rests upon well-founded scientific knowledge.

[39] This is particularly the case for those individuals that internalized an existing norm and that take a longer time to adopt to a new or changed one. See Ellickson (1999, p. 41). For the internalization of informal rules, see the "optimization"-section below.

[40] For a more general point of view, see also Posner and Rasmusen (1999, p. 380): *"[N]orm creation is too slow to provide for all the rules necessary for the governance of society [...]."*

to access personal files of other members – unquestionably influence the likelihood of such accesses taking place.

And second, there is the mechanism of internalization. Like it is the case for law-like formal rules, informal rules can also be internalized by the individual members. The underlying mechanism is basically the same: A member that internalized a rule which condemns the access to other's personal files or the use of other USB-devices than keyboards and mice would then enforce this informal rule against himself because he would otherwise feel guilty, for example.[41] Additionally, internalization of an informal rule also provides motivation for the individual members to engage in the distributed and spontaneous enforcement of that rule and thereby heightens the efficacy of the mechanism even further.

Of course, all further aspects of internalization already mentioned for the case of formal rules above apply here, too: Realizing internalization causes costs, internalization is what the frequently mentioned concepts of "awareness raising" and of establishing a "security culture" might really be aimed at, and it is unclear to what extent internalization can actually be realized in an efficient manner for the field of security-related member behavior.

Even if various aspects and research questions have to remain open as their detailed examination would go far beyond the central scope of this work, we can fundamentally characterize the use of norms-like informal rules for the purpose of hierarchical motivation by at least medium costs having to be borne for their initial creation and for any change of rule-content, by low costs arising for the centralized instance with regard to the ultimate enforcement and by a comparably high level of residual loss resulting from non-conforming member behavior.

Conclusion:
Besides architectural means and formal rules, norms-like informal rules represent the third abstract concept for realizing hierarchical motivation. Even if such informal rules have to emerge in a bottom-up direction, their development and change can nonetheless be influenced in a top-down manner. Realizing this indirect hierarchical motivation raises considerable costs for the initial creation of a constraint and for any change having to be made. The ultimate enforcement is, however, realized by the members of the community themselves and does thus not raise significant costs for the centralized instance. And finally, informal rules are of high violability and can be changed only slowly, resulting in high residual losses having to be expected.

[41] See, for example, Ellickson (1999, p. 5): *"A person who has* internalized *a norm as a result of socialization enforces the norm against himself [...]"* (emph. in original).

7.3 Motivation Costs and the Role of Information Asymmetries

After having characterized the different meta-measures and the respective motivation costs on a general basis, we will now concentrate on a specific aspect in slightly more detail that already played a significant role for the examination of the hierarchical *coordination* process in chapter 6: the effect of information asymmetries.

As constitutive property of principal-agent relations, information asymmetries strongly determine the agency costs having to be borne by the principal. The smaller the information asymmetry between principal and agent, the less incentives does the agent have to act against the stated interest of the principal because of an increasing probability of unwanted behavior being detected and because of the adequacy of agent behavior being easily assessable. With larger information asymmetries, in turn, the agent faces more incentives to behave opportunistically because of a lower probability of being caught and because of the possibility of assigning bad outcomes to adverse conditions. Generally speaking, larger information asymmetries cause larger agency costs.

What, then, are the effects of these information asymmetries for our three motivational meta-measures of architectural means, formal rules and informal rules?

For motivation being realized through architectural means, the effect of information asymmetries for the motivation process is nearly nonexistent. As those mechanisms are self-enforcing and of low violability, individual members face large or nearly infinite costs having to be borne in order to break the respective rule, independently from the level of information asymmetries currently existing between the centralized instance (the principal) and the member himself (the agent). It was already mentioned in section 7.2.1 above that we possibly do not even have a principal-agent relation anymore with architectural means being used. Be this the case or not: The existence and level of information asymmetries is scarcely relevant for the motivation costs arising from the use of architectural means, anyhow.

Take, for instance, classical access control mechanisms. If the centralized instance "motivates" a certain member through such a mechanism not to access a specific file, it makes no difference whether there are information asymmetries or not. It does not matter if the individual member sits right in front of the CISO, sharing all information with him or if the individual member resides in a hotel room and holds large amounts of exclusive, situation-specific knowledge. The motivational effect of the technical mechanism is always the same.

This is different for the meta-measure of formal rules. As enforcement is in this case realized on an ex-post basis and as formal rules are basically of a high violability, information asymmetries are of significant importance in this case. As outlined above, the individually perceived risk of rule-breaking mainly depends on the probability of rule breaking-behavior being detected and sanctioned and on the severity of the respective sanctions. Of these two factors, information asymmetries between the centralized instance and the individual member naturally affect the probability of rule-breaking behavior being detected. Larger information asymmetries lead to a lower

probability of being caught, thereby lower the individually perceived risk attributed to rule-breaking behavior, limit the motivational effect of the mechanism and ultimately induce more rule-breaking behavior. Consistently with classical agency theory, higher information asymmetries lead to higher agency costs.

A formal rule forbidding the access to a certain file or the use of other USB-devices than keyboards and mice, for example, is only effective with a certain probability of rule-breaking behavior being detected. In this case, it definitely makes a difference whether the would-be rule breaker sits right beside the CISO or in a hotel room, being aware of the asymmetry and thus knowing that nobody would actually realize rule-breaking behavior. In the hotel, the probability of being detected is significantly lower than in front of the CISO and consequently, the individually perceived risk and the resulting effect of deterrence would also be lower, leading to a higher probability of rule-breaking under conditions of higher information asymmetries.

The principal has different possibilities to work against this effect. It might, for example, be more efficient to invest in monitoring activities[42] to lessen the information asymmetries than to accept the increased amount of residual losses. The technical optimizations mentioned in section 7.2.2.1 serve exactly this goal: A notification mechanism that reports file accesses to the centralized instance reduces asymmetries, heightens the probability of detection and thereby lowers agency costs. On the other hand, such mechanisms cannot cover the respective situational information in its entirety and do furthermore cause additional costs. Even if the adverse effect of increased information asymmetries can thus be limited by such mechanisms, the overall effect is still negative. Higher information asymmetries cause higher motivation costs for the meta-measure of formal rules.

For the meta-measure of informal rules, the decisive information asymmetries are not those between the centralized instance (the principal) and the members (the agents) but rather those between the different members themselves. Due to the principle of distributed enforcement outlined above, the detection of rule-breaking behavior of one member has to be realized by other members. The higher the asymmetry between would-be rule breaker and would-be detector is, the lower is the probability of being caught and consequently, the lower is the individually perceived risk of acting against the rule. With larger asymmetries between the different members, individuals are thus less likely to follow an informal rule and more likely to behave opportunistically instead.

Take the example of an informal rule against accessing other members' personal files already stressed above. This informal rule is not enforced by a dedicated instance but rather through mechanisms of spontaneous, distributed enforcement. Any member noticing rule-breaking behavior of another one can impose different kinds of social punishment in the form of gossip, ostracism or refusals of future cooperation. Again, the underlying mechanism of deterrence mainly depends on the probability of detection and the severity of sanctions. And like it is the case for formal rules, increased asymmetries again decrease the probability of rule-breaking behavior being detected

[42] In this case, "monitoring" shall be understood in the narrow sense, confined to observation-like activities.

and ultimately result in more residual losses. A member that is strongly isolated from all other ones will thus less likely follow the informal rule not to access other members' personal files than a member who is continuously surrounded – and thus observed – by other members. Again, increased asymmetries result in higher overall costs.

Altogether, information asymmetries do not only play a relevant role for the process of coordination but rather have significant impact for the costs of motivation, too. While information asymmetries are not or only marginally relevant for architectural means, they unquestionably affect the costs of law-like and norms-like mechanisms of motivation. For motivation being realized through formal rules, higher information asymmetries between the individual member and the centralized instance induce a lower probability of rule-breaking behavior being recognized and thereby either increase the residual loss or the (broadly understood) monitoring costs having to be borne by the centralized instance. For informal rules, increasing information asymmetries among the different members lessen the efficiency of spontaneous enforcement, either leading to an increased residual loss or necessitating higher efforts of fostering internalization. Generally speaking, increasing information asymmetries thus lead to increased motivation costs – at least for the meta-measures of law-like formal and norms-like informal rules. Besides the basic characteristics of the different meta-measures, this effect also has to be considered for a close examination of hierarchical motivation cost with regard to information security.

Conclusion:
Besides the different characteristics of the different meta-measures, the level of information asymmetries influences the costs of motivation with regard to security-related member behavior, too. With motivation being realized through law-like, formal rules, hierarchical motivation costs depend on the information asymmetries between the centralized instance and the individual members. For norms-like, informal rules, hierarchical motivation costs primarily arise from asymmetries among the different members. The costs of motivation being realized through architectural means, however, are not or only marginally affected by the respective level of information asymmetries.

7.4 Coordination and Motivation – Interrelations

As we have seen so far, cooperation with regard to security-related behavior inside organizations is nowadays primarily realized through hierarchical approaches. We have also seen that economic principles can be applied to the task of hierarchical coordination as well as to that of hierarchical motivation. Coordination and motivation have, however, so far been considered separately. Even if this already provided a multitude of abstract and theoretically founded insights, the whole picture only emerges from some additional considerations regarding the interrelations between both aspects. These considerations shall be conducted in the following in order to round up the modeling of current practices and – after a short preliminary conclusion – proceed with the discussion of current and future developments subsequently.

Reconsider, for this purpose, the original model of the "three waves" of information security, on the basis of which our model of the "three eras" was developed in section 2.1. This model mainly focuses on the historical change of prevailing security mechanisms over time. Even if the model of waves was criticized for various reasons above, the repeated changes of prevailing security mechanisms did unquestionably happen. Our derived model of different "eras" of information security therefore also refers to different approaches prevailing during the different eras. The preceding considerations do now allow us to discuss this relation between different computing paradigms and different approaches to information security in a more theoretical and abstract manner.

As outlined in section 2.1.4, the first era was shaped by the paradigm of isolated computers to which only a small group of system operators had a need for access. In section 6.1, it was shown that the determination of an aspired state of security-related member behavior was comparably simple for such systems: Only few relations had to be considered and decisions could be made under virtually perfect information. This resulted in a low level of efficiency losses arising for the process of *coordinating* security-related member behavior during the first era of isolated computing.

On the other hand, the systems of this first era were, as outlined in sections 2.1.1 and 2.1.4, primarily protected by physical means. Within the more abstract understanding developed in this chapter, these physical means have to be identified as instruments for motivating members (as well as non-members) to act in conformance with the state of desired behavior that was determined during the previous coordination process. A locked door does, for example, heighten the costs of accessing the respective system for any individual that – according to the coordination outcome – "should not" access the system and that was consequently not granted physical access.

Due to the nearly nonexistent losses arising from the coordination process, physical means prove to be useful during the first era. As particular case of the architectural means characterized in section 7.2.1, physical protection raises a certain amount of initial setup costs, but once established, the costs of ultimate enforcement are low. The same is also true for the remaining losses which are, due to the low violability of physical means, virtually nonexistent. And finally, the need for repeated reconsiderations was low (see section 6.1), leading to seldom changes in the outcome of the coordination process and ultimately making the marginal costs of adaptation less relevant.[43] With physical means as the primary instrument for realizing cooperation with regard to security-related member behavior during the first era of isolated computing, the only significant costs of cooperation were thus the fixed costs of initial erection.[44] Efficiency losses arising from the process of coordination were, however, negligible.

This changed with the advent of mainframes during the second era. As outlined in section 6.2, this paradigm shift led to a higher complexity of the coordination process and thus caused higher efficiency losses because of the capacity limits of a singular,

[43] The marginal costs of providing an additional member with physical access because of a changed coordination outcome were, however, not overly high. Handing over a key and registering the respective member as key-holder, for instance, is a comparably uncomplicated task.

[44] Additional costs – even though unquestionably existing in various forms – are again omitted to simplify matters.

centralized coordination instance being reached. As also outlined in section 6.2, these efficiency losses could be minimized by means of delegation and through generalization, but in any case, a certain amount of divergence between the coordination outcome and the (hypothetical) optimal state of member behavior had to be accepted.

Regarding the aspect of motivation, technical means like access control mechanisms complemented and – to a certain extent – superseded the physical means of the first era. Different from physical means, they allowed to address the newly arising needs of isolation between different processes and users and made it possible to realize motivation on the significantly higher level of granularity already present for the coordination process. This need for a higher granularity might give an additional explanation for the switch from physical to technical means.

From an abstract point of view, technical means do, however, share a couple of properties with the formerly used physical means. As they also belong to the meta-measure of architectural means, they are also of a low violability and do thus not raise the issue of remaining losses. The costs of ultimate enforcement are effectively nonexistent and technical means also raise considerable initial costs of setup and configuration.[45] A difference between the first and the second era might, however, be seen with regard to the aggregated marginal costs of reconfiguration being necessary because of changed coordination outcomes. Such reconfigurations were – compared to the former era of isolated systems – necessary much more frequently with mainframes being used by different members at the same time.[46] Due to the repeated changes of coordination outcomes, the aggregated marginal costs of realigning the motivational means were thus not negligible anymore and had to be considered as additional cost factor.

With regard to the costs for realizing cooperation in the field of security-related member behavior, the main differences between the first and the second era thus consist in the existence of noteworthy efficiency losses for the coordination process and in additional marginal costs of motivation arising from the need for repeated reconfiguration of the motivational instruments.

While the interplay between coordination and motivation was not overly complex during the first two eras and the change of motivational instruments could directly be explained by technological change, this became different with the transition to the third era of distributed, PC-based computing. As outlined in section 6.3, this change brought with it a high level of information asymmetries and uncertainty under which the process of coordination had to take place. Even if different methods of optimization were applicable in this context, these conditions led, as a matter of principle, to coordination outcomes that strongly differed from the hypothetical optimum and were thus largely suboptimal. Would this centrally determined state of member behavior be rigidly enforced by architectural means, the organization actually had to bear the full costs of maladaption *plus* the costs of enforcing the strongly suboptimal state of aspired member behavior through technical means of motivation.

This might lead us to an additional explanation for the increasing use of formal

[45] Additional fixed costs of acquisition might be present for some technical means and nonexistent for others.

[46] See section 6.2.

and informal rules for motivational purposes besides the possibility of direct cost advantages developed in the sections above: Due to the high violability of formal and informal rules, the ultimate decision about actual behavior is – under influence of the respective individually perceived risk attributed to rule-noncompliance – made by the individual members (the agents) themselves. In this sense, the change of motivational instruments can also be seen as a mechanism of *delegating* decisions of coordination to the agents. Delegation of decisions to "lower layers" that have more complete information is, in turn, a well-established mechanism for minimizing efficiency losses of hierarchical coordination.[47] By letting the individual member decide over the use or non-use of a certain unknown USB-device, the organization is, for example, able to overcome the problem of the centralized instance not having enough information about the respective pros and cons to make efficient decisions.

Uninfluenced decision-making by the members themselves would, however, lead to selfish decisions being made and ultimately result in no cooperation being present at all. In this context, formal and informal rules give individual members an incentive to align their coordinating decisions with the collective interest of the organization as a whole. A formal rule declaring the use of other USB-devices than mice or keyboards illegitimate does, for instance, allow the individual to decide against rule-obedience when he expects no substantial punishments to be inflicted later because of the specific situation justifying the use of a certain device. Especially when concepts like "reasonability", "adequacy" or "necessity" are constituents of a formal or informal rule, this represents an act of delegating the ultimate coordination decisions to the individual members themselves. In such cases, the efficiency losses of coordination and the remaining losses of motivation do not necessarily have to be simply summed up as costs. Instead, disobedience of a rule that is itself the result of strongly imperfect coordination might very well be in the interest of the organization and lead to lessened instead of increased losses.

Formal as well as informal rules thus do not only allow to motivate individual members to act in accordance with a previously determined state of aspired behavior. They also make it possible to delegate decisions of coordination to the individual member and to give this individual member an incentive to decide in the collective interest of the whole organization. The switch from architectural means to formal and informal rules thus must not be seen in the light of motivation alone but also represents a mechanism for minimizing *coordination* costs. Even if necessitating substantial amounts of motivation costs to be accepted by the organization, formal and informal rules thereby allow to overcome – to a certain extent – the substantial costs of coordination arising from the increased complexity, information asymmetry and flexibility that shape the third era of decentralized, PC-based computing. Besides the possible advantages in matters of pure motivation costs mentioned in section 7.2, this might provide an additional explanation for the change from architectural means to formal and informal rules that happened with the transition from the second to the third era.

[47] See section 3.2.1 for the general case and sections 6.2.1 and 6.3.1 for a more explicit focus on information security.

Conclusion:
Even if coordination and motivation can to a large extent be analyzed separately, there also exist interrelations between them. These interrelations provide additional explanations for the repeated shifts of prevailing motivational mechanisms that happened together with the shifts of prevailing computing paradigms. In particular, the shift from physical to technical means can, besides other reasons, be explained by the need for higher granularity of coordination resulting from the shift from the first to the second era. The shift from the technical means being used during the second era to the formal and informal rules that were added during the third era can be explained by an economic need for delegation of security-related decisions to individual members. In this model, formal and informal rules do not only provide individual members with incentives to follow some kind of determined coordination outcome but also motivate them to make coordination decisions that follow the collective interest of the whole organization.

7.5 Preliminary Conclusion and Further Issues

In this chapter, we concentrated on the hierarchical process of *motivation* with regard to security-related member behavior inside organizations. We reconsidered the three meta-measures of architectural means, formal rules and informal rules already identified and initially characterized in section 2.2 and showed that they are equivalent to the well-known modalities of architecture, law, and social norms as introduced by Lessig. This correspondence, together with the established economic model of principal-agent relations allowed us to develop an economic model of costs arising from the different meta-measures being used for hierarchical motivation.

Based on this model, we discussed the different cost structures of the different meta-measures. In particular, we distinguished the initial, fixed costs of setting up the respective motivational instrument which have to be borne by the centralized instance only once, the marginal costs of realigning the instrument in case of a changed outcome of the coordination process, and the ultimate enforcement costs that have to be borne for any single case of motivating the member to behave in conformance with the coordination outcome. Additionally, we also examined the residual losses arising from rule-breaking member behavior for any of the three meta-measures.

We have shown that the three meta-measures feature strongly different cost structures. Architectural means are characterized by significant initial setup costs and a medium level of marginal costs arising for the case of altered coordination outcomes. On the other hand, their self-enforcing nature leads to no or only negligible costs having to be borne for any single case of motivating a certain behavior and the ex-ante approach of enforcement leads to a low or non-existent level of residual losses. Formal, law-like rules, in turn, cause low fixed costs of initial set-up, a medium level of costs for adaptation to changed coordination outcomes and comparably high costs arising for any single case of rule enforcement. Due to the ex-post enforcement model, formal rules are of a high violability and thus cause a medium level of residual loss. And

finally, norms-like, informal rules raise medium costs of initial setup, medium to high marginal costs of repeated adaptation, and – from the perspective of the centralized instance – low enforcement costs for any single act of motivation. The residual loss, however, has to be considered as high. These different cost structures are aggregated in table 7.6.

Table 7.6: Costs of hierarchical motivation – Summary.

Meta-Measure	Fixed Costs	Marginal Costs	Enforcement Costs (single case)	Residual Loss
Architectural means	high	medium	none / negligible	none / negligible
Formal Rules	low	medium	high	medium
Informal Rules	medium	medium – high	low	high

Especially for the meta-measures of formal and informal rules, some possibilities of optimization were identified. In both cases, internalization of the respective rule by the individual member can heighten the individual "costs" being associated with rule-breaking behavior. If an individual member has internalized a certain rule, he would, for example, feel guilty if he did not obey the rule and this expected feeling of guilt affects the individual payoff function of rule-breaking behavior, ultimately leading to a higher level of rule-obedience. Furthermore, internalization also affects the effort that members are willing to invest in the spontaneous enforcement of informal rules. It would, however, go far beyond the scope of this thesis to estimate the extent to which internalization represents a viable mechanism of cost optimization. This question has therefore to remain open.

As a second approach of optimization, we identified the possibility of supporting the enforcement of formal rules by technical means. Various kinds of monitoring systems could, for example, heighten the probability of rule-breaking behavior to be recognized by the centralized instance, thereby heighten the individually perceived risk of disobedience and ultimately lower either the residual loss or – as a result of realigned motivational efforts – the enforcement costs. Besides their direct application as motivational instrument, architectural means can thus also be used in an indirect manner to reduce the enforcement costs of formal rules, for example.

Furthermore, the effect of different levels of information asymmetries on the motivation costs arising for each of the three meta-measures was examined. Generally speaking, an increasing level of information asymmetry between the individual member and the centralized instance leads to higher costs for formal rules and increasing asymmetries between different members make the mechanism of informal rules less efficient while the costs of architectural means are not significantly affected by the

existing level of information asymmetries.

As a final aspect, we discussed some interrelations between coordination and motivation. In particular, we showed that the change from architectural means to an increasing use formal and informal rules during the third era might also be understood as an act of delegating coordination decisions to the individual members themselves in order to lessen the efficiency losses that would otherwise have emerged from centralized coordination. Instead of merely motivating members to behave in accordance with some previously determined state of aspired behavior, formal and informal rules do in this case also motivate the individual member to make decisions that reflect the collective interest of the organization as a whole.

Again, all these conclusions shall be understood in an abstract, broad and generalized sense that explains the *basic* economic characteristics of the different meta-measures being used for the purpose of motivation with regard to security-related member behavior. It is, nonetheless, possible to derive some preliminary explanations and implications from the findings of this chapter.

The remarks on the different cost structures, for instance, might provide an explanation for the widely known phenomenon of voluminous security guidelines (formal rules) being formulated, enacted and constantly extended while at the same time not being vehemently enforced.[48] Different from architectural means and informal rules, their initial setup does not raise reasonable costs and does thus not require significant justifications. The ultimate enforcement of such formal guidelines, however, would require substantial efforts to be made and the respective high costs might give an explanation for organizations abstaining from this final step in many cases.[49]

Another implication regards the different possibilities of optimization outlined above. Even if not consciously aimed at the abstract concept of internalization, a wide variety of established activities usually conducted under labels like "user trainings" or "awareness campaigns", for example, implicitly pursues the goal of making individual members internalize the respective formal and informal rules. Being aware of the applicability of the abstract concept of internalization, the respective activities can thus be organized much more consciously. Applying the existing theory of internalization to user trainings, awareness campaigns, etc. could, for example, help to put such motivational activities on a scientific, theoretical basis and to raise their effectiveness and efficiency. Again, this opens a wide field for future research.

And finally, the remarks on the influence of information asymmetries might be of high importance to the whole field of information security within highly distributed and mobile settings. At least, the outlined effects raise doubts in the sustainability of existing law- and norms-like motivational approaches for organizations that mainly consist of locally dispersed, autonomously acting members. This issue will, however, be considered in more detail in chapter 9.

Further issues mainly arise from the economic field of agency theory where the problem of adverse selection with regard to the choice of agents is an integral aspect:

[48] See, for example, Shostack and Stewart (2008, p. 129).

[49] Such an explanation would also be backed by Cheng's (2006, pp. 668 f) more general notion of *"institutional pressures"*, which might provide a stimulating starting point for future discussions on this issue.

With substantial information asymmetries existing between a principal and a multitude of would-be agents, the principal can hardly make the right decision which agent to choose, ultimately resulting in a lemons-market with only low-quality agents remaining.[50] Mapped to our more specific subject of organization-internal information security, one could ask how an organization could be enabled to incorporate security-related aspects into their choice of members – not only with regard to the initial admission of new members but also for the choice among existing members for a certain project, for example. Of course, there are a multitude of other factors typically being more relevant for such decisions than those related to information security, but in some cases, a member's ability and willingness to behave secure might be decisive. It might thus be interesting to reconsider well-established approaches to the problem of adverse agent selection – like the concepts of screening[51] or that of meaningful signals developed by Spence (1973)[52] – in the light of security-related behavior that could be expected from various would-be members of an organization.

Another well-known and widely discussed problem from agency theory is the risk of moral hazard – the opportunity for an agent to ascribe good outcomes to high efforts and bad outcomes to adverse conditions already mentioned above and the fact of agents taking unreasonable risks because of not having to bear the costs of negative results.[53] To overcome the problem of moral hazard, agency theory suggests at least a certain amount of risk to be held by the agent in order to provide him with incentives to incorporate risks appropriately.[54] For the field of security-related member behavior, such risk-transfer to the agent could, for instance, be realized through individual liability for the case of incidents occurring on the side of an individual member.

As we can see, several problems of information security are on an abstract level widely established within general agency theory. The application of existing research from agency theory would thus presumably be a rich source for conducting theory-based research in the field of information security – especially with regard to the motivation of security-related member behavior. This would, however, again go far beyond the primary scope of this thesis and shall thus only serve as suggestion for future research activities.

[50] For the general concept of lemons-markets, see, above all, Akerlof (1970).

[51] Interestingly, the ISO standard 17799 also mentions *"screening"* as a part of human resources security (see ISO / IEC 2005a, section 8.1.2).

[52] In this model, it is necessary for signaling to be an adequate and reliable instrument that *"the costs of signaling are negatively correlated with productive capability"* (id., p. 358) of the agent to ensure that only those agents with high productivity take the signaling costs. Note, however, that *"the negative correlation is a necessary but not sufficient condition for signaling to take place"* (Spence 1973, p. 367).

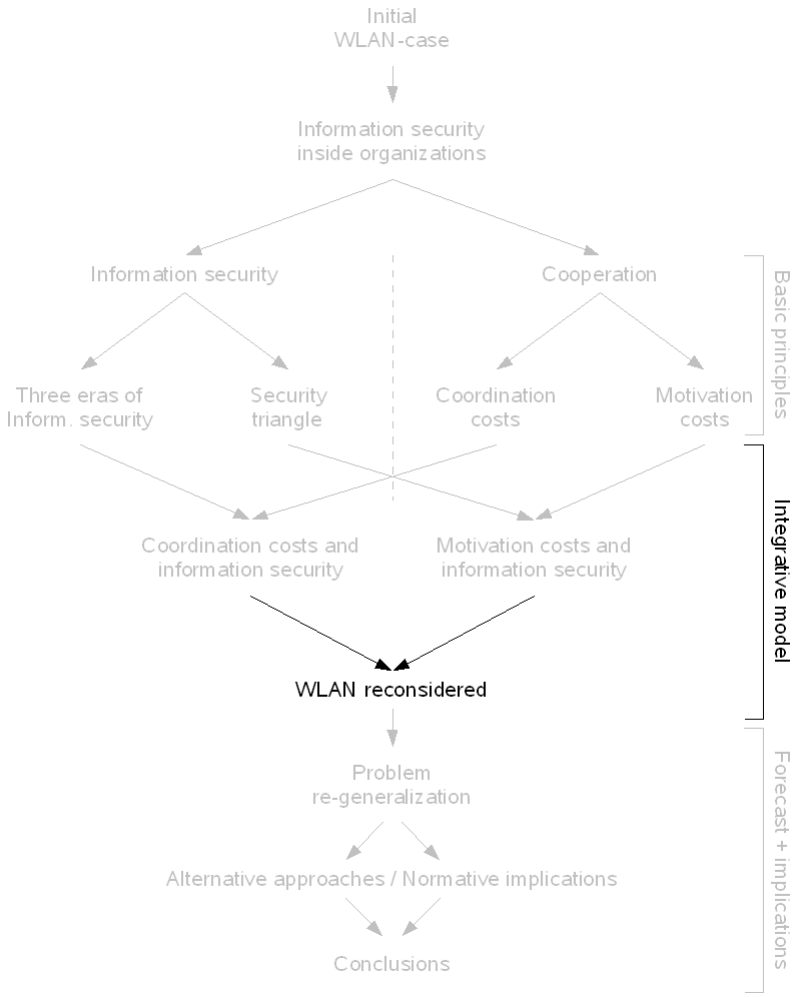
[53] Anderson (2008, pp. 823 f) also uses the term *"moral hazard"* but does so without reference to – or trying to profit from – agency theory.

[54] See, for example, Laffont and Martimort (2001, pp. 145 ff) or Milgrom and Roberts (1992, pp. 206 ff).

Conclusion:
Like for any other kind of hierarchical motivation, the centralized instance and the individual members are in a principal-agent relationship for the case of security-related member behavior, too. The necessary incentives for individual members to behave in conformance with the outcome of the coordination process can be created through architectural means, law-like formal rules and norms-like informal rules by the centralized instance. These three meta-measures feature different cost structures. In particular, architectural means raise high initial costs but low costs of ultimate enforcement and cause low residual losses. Formal and informal rules are cheaper to create but cause higher residual losses and formal rules raise considerable additional costs of ultimate enforcement. Furthermore, formal as well as informal rules are strongly affected in their effectiveness by the level of information asymmetries being present. The higher these asymmetries are, the higher are the motivation costs arising from these two meta-measures.

Chapter 8

The Use of Public WLAN – Reconsidered



Chapter 8

The Use of Public WLAN – Reconsidered

And here, poor fool! with all my lore
I stand, no wiser than before

– Johann Wolfgang von Goethe,
Faust: The First Part of the Tragedy

We shall now return to our original and more tangible problem of members of an organization using publicly available WLAN-hotspots. As it already became clear from our initial remarks in section 1.2, information security plays a significant but at the same time nontrivial role in this context. We will thus reconsider the use of public WLAN on the basis of the abstract concepts developed so far.

It was outlined in section 5.1 that the use of public WLAN by one member of an organization can result in strong negative externalities for all other members and that WLAN usage thus requires cooperation to be realized among individual members. Like it is the case for any other subject of cooperation, this necessitates *coordination* to determine the state of member behavior that would represent the highest payoff for the organization and *motivation* to make the members behave in accordance with the previously determined "optimal" state.

For the context of public WLAN, the process of coordination refers to the question whether a certain member of an organization "should" or "should not" access a certain publicly usable hotspot in a specific situation. Giving the right answer to this question from a collective perspective – that is, realizing perfect coordination – would require a multitude of factors to be taken into account that shall be addressed in brief to proceed with the analysis of cooperation costs subsequently.

8.1 Specifics of the WLAN-Case

Besides the general complexity that already arises from the fact of PCs being used within WLAN-based settings – remember the remarks on coordination costs during the third era in section 6.3 above – perfect coordination would in principle require *any* public WLAN hotspot to be taken into account as additional entity. Like it was the case for the entities that had to be added to the coordination process and increased complexity and uncertainties during the third era, these additional entities

are principally unknown to a centralized coordination instance and increase complexity and uncertainties even further. Publicly usable WLAN-hotspots could thus be seen as just another class of unknown entities that perfectly fit into the model of hierarchical cooperation developed above.

There is, however, a difference between WLAN-hotspots and, say, USB-devices with storage capabilities: While the use of USB-devices and other unknown entities could primarily be considered to occur *within* the well-defined and well-known boundaries of the organization-internal network, the use of public WLAN hotspots constitutes a breaking of just these boundaries. Without any further mechanisms being present, notebooks that are primarily considered as part of the organization-internal network do, when connected to a hotspot, also become part of another network – be it the one of a hotel or the network of a specialized hotspot provider. Instead of merely concentrating on potentially unknown entities being connected to a specified network, a "perfect" coordination process had to take this fact into account and consider potentially unknown entities (notebooks) being connected to unknown networks other than the organization-internal one, too. Within the model of structurally different computing paradigms developed in section 2.1, this change can be illustrated as done in figure 8.1.

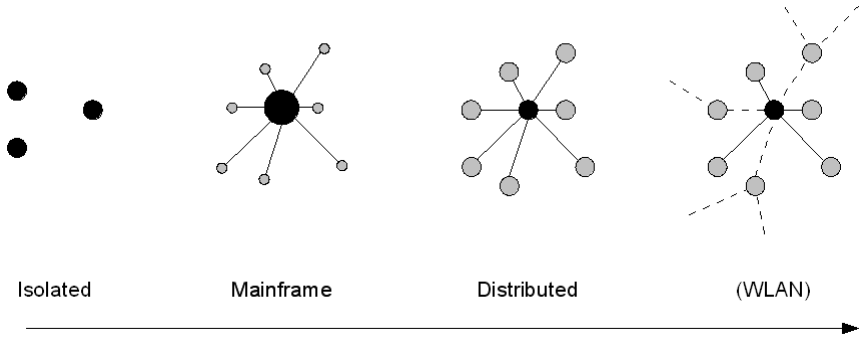


Figure 8.1: WLAN-usage and the three eras of computing

Besides the fact of external, potentially unknown and generally insecure networks having to be taken into account, we also have to consider the fact that for typical scenarios of hotspot usage, the specific situational context is much more relevant for information security than it is the case for traditional office environments. Instead of being able to make decisions on the basis of well-known and comparably constant physical conditions, the local environment now also has to be considered as ever-changing and principally unknown to a centralized instance. And as this environment does at least influence the risk arising from hotspot usage[1], this increases the uncertainty for

[1] It might, for example, make a difference with regard to the threat of shoulder-surfing whether the member resides in the lobby of a hotel or two floors above in his personal room while accessing a hotspot and entering a VPN-password. Other factors possibly influencing the situation-specific risk include further location-dependencies (country, city, etc.) or even the time of day or the

the coordination process even further.

At least similarly important as the situation-specificity of risks is the fact that the *benefits* possibly being realizable through hotspot usage can also not be known in advance. Generally speaking, it is in many cases just the possibility to react flexibly to situation-specific and unforeseeable needs that constitutes the mobility of members and thus the use of hotspots at all. While organizations are on the one hand able to generate profits from such situation-dependent, ad-hoc responses to unforeseen conditions and requirements, this does on the other hand imply additional uncertainties about the appropriateness of hotspot usage in a given situation. If, for instance, a mobile member wants to access a hotspot on an airport, it cannot be known by a centralized instance in advance whether he wants to do so in order to simply shorten the time of waiting for a flight or whether he wants to access the organization-internal customer database to serve an immediate client-call. Even for such trivial tasks as accessing the organization's email server, the trade-off might strongly depend on the respective situation and require context-specific information to be taken into account that can, as a matter of principle, not be known by a centralized coordination instance.

Both specific aspects, the need to consider unknown, organization-external and possibly insecure networks on the one hand and the strong situation-dependence of risks and possible benefits on the other, increase information asymmetries and uncertainties for the coordination process even further. Organizations have to respond to these facts when trying to realize cooperation among members within the context of hotspot usage. As we will see in the following, the three intuitive solutions from section 1.3 represent three different approaches to do so.

8.2 Intuitive Approaches and Costs of Cooperation

After these preliminary remarks, we can now reassess the three intuitive approaches initially outlined in section 1.3 within our framework of cooperation costs developed so far. Remember that the three approaches were distinguished as follows:

- A *general ban* realizes strong isolation between the organization-internal network and any "outside"-entities like publicly usable hotspots. Typically, this approach includes the application of technical solutions like "enforced VPNs" that prevent logging on to public WLAN hotspots through the browser-based Universal Access Method (UAM).

- *Login automation* refers to technical solutions that perform the logon-process to a hotspot automatically and without user interaction. Instead of the UAM, a separate and usually provider-specific software does in this case perform the authentication process. Additionally, this software is typically combined with a VPN-client that automatically establishes a secure connection to the organization-internal network.

number of other active WLAN devices in the same area.

- A *lax treatment* was considered as third possible option for addressing the problem of hotspot-usage. This approach is less strict and basically accepts insecure connections on higher network layers in order to allow members to log on to a hotspot via the Universal Access Method. Members could, however, be instructed through corresponding prescriptions to manually establish a VPN as soon as possible.

From each of these three approaches, we can extrapolate to the respective underlying model of coordination and identify the abstract mechanisms being used for motivation purposes. Based on these extrapolations, we can then analyze the approach-specific cost structures and compare them with each other.

8.2.1 General Ban

The first approach of a general ban resembles the initial use of PCs as mainframe-terminals in the beginning of the third era. Instead of really responding to the new conditions, the previous state is artificially recreated – or retained – in order to prevent the increased complexity from gaining relevance at all. By denying *any* access to publicly usable WLAN hotspots, a rise of complexity is prevented and instead of actually responding to the newly arising challenge, the challenge itself is eliminated. This allows to maintain established and well-known practices and does not result in any additional losses having to be borne.

On the other hand, such an elimination also prevents any benefits that could otherwise be derived from the changed conditions. Even if the use of WLAN would unquestionably be in the very interest of the organization, the coordination process would nonetheless come to the conclusion that hotspots "should not" be accessed, thereby leading to coordination costs in the form of maladaption. Instead of representing an elimination of the new challenge, a general ban could thus also be interpreted as *overgeneralization* that results in considerable efficiency losses in the coordination process. The coordination process is then basically the same as it was without the possibility of hotspot-usage. No additional coordination costs arise as a result of increased complexity and uncertainty but a considerable risk of maladaption as a result of suboptimal coordination outcomes does very well exist.

Motivation is, in turn, realized through architectural means of low complexity. The only "motivational" task having to be accomplished is the automatic blocking of any higher-level connection to unknown networks. Usual solutions realizing an "enforced VPN", for example, do exactly this. Like it is the case for other architectural means, the use of such solutions requires some initial costs of acquisition and configuration to be borne but once installed, they realize any single act of motivation – that is, in this case, the prevention of connecting to a hotspot – at effectively nonexistent costs. As individual members can only circumvent such solutions at substantial individual costs, no or only neglegible residual losses have to be expected. And as the outcome of the coordination process is not likely to change at all, marginal costs of reconfiguration can be neglected, too.[2]

[2] In section 1.3.1, we also mentioned "acceptable use policies" to be commonly used together with

Altogether, the first approach of a general ban can thus be characterized by a coordination process that does not respond to the new challenge posed by publicly usable hotspots but rather eliminates the challenge itself by generalizing over all hotspots and all situations, thereby leading to the risk of substantial costs of maladaption having to be borne by the organization. Motivation is in this approach realized by architectural means that raise a certain amount of initial fixed costs but result in no or only marginal costs of reconfiguration and of ultimate enforcement. Residual losses also do not have to be expected.

8.2.2 Login Automation

For the second approach of login automation, the situation is comparable but not similar. Like for the case of a general ban, the underlying coordination process leading to such a solution must basically have considered the entirety of all hotspots in a uniform, generalized manner and must have come to the conclusion that access "should" only be realized when any higher-level connection is protected by a VPN. But different from the highly strict strategy represented by the general ban, the approach of realizing an automated login procedure makes the concession of a short higher-level connection being established with an external network in order to allow hotspot-usage at all. Instead of trying to eliminate the challenge itself, the changed conditions are thus considered more consciously.

The coordination process must then have concluded that the risks arising from a short higher-level connection that is independent from user-interaction are justified by a higher amount of benefits being derivable from hotspot usage – even if the number of hotspots is, as outlined in section 1.3.3, still strongly limited – and that the risks should thus be accepted. Different from the change-denying approach of a general ban, the approach of login automation is thus a sign for a real weighting between risks and benefits to have taken place. Due to this more conscious coordination, the risk of losses arising from maladaption is at least lower than for the approach of a general ban.

Motivation, however, is in this case realized like it is for the general ban. Architectural means ensure that an unprotected higher-level connection is solely used for authentication purposes and prevent any kind of user intervention. Like with the general ban, any single member is hindered from acting against the collective interest of the organization through technical means. These technical means do again cause a certain amount of initial fixed costs but raise no noteworthy costs of reconfiguration and cause no residual losses. No member can, for instance, circumvent the organization's firewall and introduce malware into the protected internal network or use an unprotected connection to export large amounts of internal information. Even if a certain

technical means. As, however, the technical means do not raise considerable losses as a result of non-conforming behavior, such law-like policies are at least not needed to lower losses through a heightening of the individually perceived risk associated to rule-breaking. An explanation for such policies being nonetheless established might be the creation of a law-like legitimation for the architectural means to be used. In this case, the enforcement costs of the law-like formal rules are irrelevant.

amount of risk is thus accepted within the approach of login automation, members still have no noteworthy opportunity to act in nonconformance with the state of behavior determined during the coordination process. Like for the first case of a general ban, motivation costs are thus mainly limited to the initial, fixed costs of acquisition and setup, while coordination costs are higher for the coordination activity itself but lower with regard to the risk of suboptimal coordination outcomes.

8.2.3 Lax Treatment

The third approach of a lax treatment, in turn, strongly differs from the aforementioned two. Most obvious is the difference in the *motivational* mechanisms being used. Instead of enforcing some state of desired member behavior through architectural means, motivation is in this case realized through formal and informal rules. As outlined in sections 7.2.2 and 7.2.3, both of these meta-measures function on an ex-post basis and are of a high violability. For the context of hotspot usage, this implies that individual members are not technically prevented from establishing unprotected higher-level network connections but are rather threatened with ex-post sanctions for the case of doing so.

From a cost perspective, motivation through such formal rules tendentially causes lower initial costs than architectural means but on the other hand raises considerable costs of ultimate enforcement for any single case and, not to forget, residual losses as a result of the high violability. Informal rules, in turn, do not cause costs of ultimate enforcement for the centralized instance[3] but again carry the risk of residual losses for the organization resulting from opportunistic member behavior.

As outlined in section 7.3, all these costs strongly depend on the level of information asymmetries existing between the central instance and the individual member that should be motivated (for formal rules) or on the level of asymmetries existing between different members (for the case of informal rules). High information asymmetries are in turn integral to nearly *any* scenario that involves access to publicly usable WLAN hotspots. A typical situation for hotspot usage would, for instance, be a mobile member of the organization using his notebook on an airport, in a hotel room or even in a city park. Compared with traditional office settings, such conditions either make the discovery of rule-breaking behavior more expensive for the central instance or lead to a lowered probability of rule-breaking behavior being discovered. The probability of violations of informal rules being recognized by other members is also decreased and ultimately, the overall motivation costs are comparably high with formal and informal rules being used for the purpose of motivating security-related member behavior with regard to hotspot usage.

We will return to this issue in more detail later but for the time being, the question arises why an organization should apply this comparably costly approach to motivate members in the context of publicly usable WLAN at all? As the approach of a lax treatment is widely used by a multitude of organizations[4], there must be reasons

[3] Remember the concept of spontaneous enforcement outlined in section 7.2.3.

[4] Were this not the case, no hotspots other than those operated by large providers that offer automated login solutions would be used in an organizational context. This is obviously not the case.

to do so. These reasons might be found in the interrelations between motivation and coordination outlined in section 7.4: Different from architectural means, formal and informal rules can be used to delegate parts of the *coordination* process to the individual members themselves.

If, for instance, it has to be decided whether a certain mobile member "should" or "should not" access a hotspot in a specific situation, this decision basically requires a trade-off to be made between the risks that would be induced for the whole organization by this member using the hotspot on the one and the value that could be derived for the organization from hotspot usage on the other hand. As mentioned above, this trade-off might in turn require situational, context-specific information to be taken into account that can – due to the strong information asymmetries being present for mobile settings – not or only hardly be known by a centralized coordination instance.

Instead of simply accepting these asymmetries and making generalized decisions that cause costs of maladaption – like it is done within the first two approaches – a lax treatment allows organizations to delegate the ultimate coordination decisions to the mobile member himself in order to permit context-specific information to be included in the respective considerations at all. As also outlined in section 7.4, the member must then be motivated to make coordination decisions that actually represent the collective interest of the organization instead of making opportunistic decisions that maximize his individual payoff.[5] It is this motivation that established formal rules with regard to hotspot-usage are (at least also) aimed at: They define what constitutes – from the collective point of view – an "acceptable use" on an abstract level and then let the individual member decide whether his specific situation meets the requirements for the use of a hotspot to be "acceptable". Of course, this approach does again raise additional costs of motivation – the members' decisions must be evaluated ex-post, sanctions must possibly be exerted and a certain residual loss must be expected as a result of opportunistic member decisions[6] – but basically, the reduction of maladaption costs can over-outweigh the increase of motivation costs and thus justify the approach of a "lax treatment" for certain organizational settings.

8.3 Comparison of Cost Structures and Implications

We can thus recapitulate the different cost structures of the three different approaches to the use of WLAN hotspots in a highly generalized manner as follows: A general ban requires no coordination activities with regard to the specific context of hotspot usage but results in a high risk of maladaption costs because of WLAN not being used even when it would be valuable for the organization. Motivation costs do, due to the architectural means that are used in this approach, primarily arise as initial fixed costs but not with regard to ultimate enforcement. Residual losses arising from

Consequently, there must be organizations following the approach of a lax treatment based upon formal and informal rules instead of architectural means.

[5] We thus have a classical principal-agent relation between the organization and the mobile member for the case of a lax treatment. See again section 3.2.2 and the introduction of chapter 7.

[6] Of course, all the optimizations mentioned in sections 7.2.2.1 and 7.2.3.1 can be applied here, too. Recall, in particular, the concept of internalization.

member behavior not conforming to the coordination outcome also do not have to be expected.

With login automation, a certain risk is accepted in exchange for being able to realize the benefits that are possible through the use of a limited number of hotspots. Due to this trade-off being made, the risk of maladaption is, compared with the general ban, reduced. Motivation costs, however, are basically similar as architectural means are again used for this purpose: Some fixed, initial costs have to be borne but the ultimate enforcement does not cause any additional costs. Remaining losses are unlikely, too.

The third approach of a lax treatment, in turn, minimizes efficiency losses of maladaption by delegating the ultimate coordination decision to the individual member, who is much more able to take context-specific aspects into account than any centralized instance could at reasonable cost. The member himself must nonetheless be motivated to make decisions that actually lie in the collective interest of the organization. This requires formal as well as informal rules to be used which in turn cause significant costs for each single act of motivation and will also lead to certain remaining losses (see table 8.1).

Table 8.1: Different approaches for WLAN-security – Expectable cost structures (strongly generalized)

	General Ban	Login Automation	Lax Treatment
Coordination Costs / Risk of Maladaption	high	medium – high	low
Fixed Motivation Costs	medium	medium	low
Ultimate Enforcement	none	none	high
Residual Loss	none – low	none – low	medium – high

When, then, should which of the three approaches be chosen by an organization for realizing cooperative member behavior with regard to the use of WLAN hotspots? Of course, this decision strongly depends on the kind of organization being considered. The cost structure developed so far and summarized in table 8.1 can only represent strongly generalized principles that distinguish the different approaches on an abstract level. To utilize these principles for actual decision-making within a specific organization, they must always be considered in the light of the organization's specific givens. The consideration of two idealized example organizations shall illustrate this relation more vividly.

8.3.1 Example 1: Well-Defined Tasks and Situation-Independence

As a first example, consider a large organization with well-defined and strongly formalized tasks to be carried out by the individual members and with many of these tasks having to be conducted in non-stationary, mobile settings. Assume furthermore

that this organization could only limitedly profit from hotspot usage and that it has strong security requirements with even a small probability of adverse events leading to a high risk because of enormous losses for the case of such an event actually occurring. The field service of an insurance company might be an ostensive example for such a setting.

Even if this constructed example does – implicitly – already suggests at least a non-lax approach to be chosen, let us nonetheless analyze it with regard to the cost structures developed above for exemplary purposes: Even if there is a need for individual members to perform their task within decentralized, mobile settings, they nonetheless only have to act within a well-defined and formalized set of behavioral options and have no need for flexibly responding to unforeseen, situation-specific circumstances. Due to this task-profile, the effective complexity of the coordination process is comparably low: All relevant members as well as their (well-known and well-defined) requirements can be treated in a generalized manner and without having to take situation-specific aspects into account. Furthermore, the risk of unrealized possibilities of value-generation because of a potentially too high level of security is comparably small while the risks arising from a possibly too low level of security would be comparably high.

Within such a setting, the costs of realizing cooperation through a "general ban" turn out to be as follows: As already outlined, explicit coordination activities are only required on a nonexistent or very low level, avoiding any substantial costs or efficiency losses. The "high" risk of maladaption specified for the idealized case of a general ban above has to be relativized because of the low possible impact of a too high level of security. The respective maladaption costs of a general ban have therefore to be considered as low. The fixed motivation costs of setting up a technical solution that enforces the general ban are – consistently with the generalized cost structure – on a medium level and the costs arising for ultimate enforcement and as residual losses are also effectively nonexistent.

The approach of an automated login would lead to comparable costs. Due to the effectively similar coordination process – members can easily be generalized and coordination can be realized with respect to a well-known and well-defined set of activities – only few coordination efforts are necessary. Either all mobile members "should" use hotspots or no mobile member should. Maladaption costs, which arise from the risk of a suboptimal state of member behavior being the outcome of the coordination process, do thus play no significant role, too. If being changed at all, the costs of maladaption could be slightly lower than for the general ban if hotspot access represented a small but existing benefit to the organization that supersedes the risk being associated with the short higher-level connection to an external network. The level of fixed motivation costs as well as the costs of ultimate enforcement and the residual losses can again be considered as being basically the same as for the general ban. The overall costs of the "automated login" thus differ from those of the "general ban" only in their slightly increased coordination effort and in the possibility of the costs of maladaption being slightly reduced.

A "lax treatment", however, would induce a strongly different cost structure for the organization considered here. As the coordination effort – the determination of

an "optimal" state of member behavior – would again be basically the same as for the general ban and for login automation, the risk of a strongly suboptimal state being determined in a centralized process would still be low.[7] This would not justify a delegation of ultimate decisions to the individual member to reduce coordination costs in the sense of section 7.4. Compared to the first two approaches, our example organization would thus not be able to benefit from the use of formal or informal rules instead of architectural means but nonetheless had to bear the respective costs of motivation. These would include the costs of ultimate enforcement which are, as outlined above, substantial for typical scenarios of potential WLAN usage because of the strong information asymmetries existing between different members as well as between the centralized instance and the individual members. Furthermore, the organization had to bear a certain amount of residual losses arising from member behavior that does not conform with the coordination outcome. As we assumed a possibility of enormous losses materializing for the organization in the case of an adverse event actually occuring, the impact of such nonconforming behavior would be certainly high. While our example organization would thus not be able to profit from the delegation of coordination decisions to individual members substantially, it had to bear significant costs arising from nontechnical approaches of motivation for the case of a "lax treatment" (see table 8.2).

Table 8.2: Costs of different approaches for WLAN-security – Example 1

	General Ban	Login Automation	Lax Treatment
Coordination Effort	none – low	low	low
Risk of Maladaption	low	low	low
Fixed Motivation Costs	medium	medium	low
Ultimate Enforcement	none	none	high
Residual Loss	none – low	none – low	high

Based on these considerations, it is obvious that our imaginary example organization should at least *not* use the approach of a lax treatment. No benefits could be derived from the delegation of coordination decisions but substantial costs of motivation, including a considerable amount of residual losses, had to be accepted. Instead, the organization should rather use the approach of a general ban or that of login automation which both raise significantly lower costs. The decision between those two approaches should be made depending on the relation between possible benefits being derivable from using a limited number of hotspots on the one hand and the additional costs that would be introduced through the short higher-level network connection with

[7] In fact, costs of maladaption could even be slightly lower than for the case of login automation because of nearly *all* hotspots being usable. But as the organization would, according to our assumptions, not significantly profit from hotspot usage at all, this cost reduction is marginal as compared to the cost increases mentioned in the following.

the approach of an automated login on the other.

8.3.2 Example 2: Flexibility, Complexity and Relevance of Context

As a counter-example to the idealized organization described above, consider a smaller organization with completely different givens and requirements. Assume that most of the organization's members are classical knowledge workers having to solve non-structured tasks and that large parts of these tasks must necessarily be carried out within mobile settings. Assume furthermore that these activities induce strongly situation-specific requirements which can not even nearly be foreseen by a centralized instance and that mobile connectivity of individual members generally allows the organization to derive considerable benefits while the impact of an adverse event would be significantly smaller than for the organization from the first example. A unit of a university, a consulting firm or several kinds of creative agencies might serve as concrete representations for such an idealized organization.

A "general ban" would obviously not meet the specific requirements of this organization. The over-generalized coordination process resulting in the conclusion of no hotspot usage being acceptable would obviously not represent the optimal state of member behavior. As the use of hotspots would – even with a certain amount of risks being present – for many cases represent an overall benefit, we had to consider substantial maladaption losses having to be borne by the organization.[8] Even if the motivation costs for enforcing this strongly suboptimal coordination outcome are again effectively limited to some initial, fixed setup costs, the overall losses arising from a general ban would in this case be significant.

With an "automated login" being used, the organization would at least be able to reduce the costs arising from a strongly inadequate coordination outcome. Instead of preventing any use of hotspots, members could be determined to "should" use at least a certain proportion of them to generate benefit for the whole organization. But at the same time, the coordination process would still lead to significantly suboptimal outcomes. As mobile connectivity is assumed to be an essential basis for value generation, the fact of a multitude of hotspots not being usable would still indicate a coordination process that strongly differs from the hypothetical "optimal" state of member behavior. Even if the costs of maladaption can thus be decreased by the use of login automation instead of a general ban, they nonetheless still exist on a considerable level. The underlying coordination process, however, would again require only marginal efforts and the costs of motivation would also be similar to those from the first example.

The approach of "lax treatment", in turn, seems to meet the specific requirements of our second example organization much better. As the coordination process would in

[8] Instead, one could also assume the "general ban" approach to basically allow *two* possible outcomes: the general ban being used or no measures being taken at all. The coordination process could then also come to the conclusion that hotspots "should" be used for all cases. This would, however, also represent a suboptimal state of member behavior – members could also use a hotspot in a completely unprotected manner even if doing so were not in the collective interest – and result in certain costs of maladaption, too.

this case actually require a multitude of situation- and context-specific factors to be taken into account and as these factors can actually not be known by the centralized instance within this example, the coordination model of the two alternative approaches *must* have resulted in strongly suboptimal outcomes.[9] This inefficiency of a centralized coordination process gives reason for the delegation mechanism from section 7.4 to be applied here: Instead of trying to determine a desired state of member behavior for *any single case* of possible hotspot usage in a centralized process, the ultimate coordination decision is delegated to the individual member who is motivated by formal and informal rules to decide in the collective interest of the organization. In this case, the necessary coordination effort is reduced and the risk of strongly suboptimal decisions being made is also lessened because of the individual member being much more able to decide whether the use of a hotspot within a highly specific situation lies in the collective interest or not.

The formal and informal rules being employed for motivating the member to decide in the collective interest do, in turn, cause only few initial costs but high costs of ultimate enforcement. If, for instance, there is an "acceptable use" policy defining hotspot usage as acceptable as long as this is done "for work-related purposes" and as long as "a VPN is established as soon as possible", compliance with this policy has to be monitored and detected noncompliance needs to be sanctioned. As the efficiency of these (broadly understood) monitoring activities is lessened by the strong asymmetries and uncertainties, we essentially had to consider the respective costs as being on a high level. But on the other hand, there are various mechanisms of optimization[10] that can be applied and that could possibly relativize the impact of ultimate enforcement costs. And finally, there will – depending on the individually perceived risk of noncompliance resulting from the motivational instruments – be a certain degree of residual noncompliance that also has to be considered as costs arising from the approach of a lax treatment. These costs can, however, be assumed to be lower than for the first example organization because of the smaller impact of adverse events that we assumed above.

While a general ban as well as login automation thus *must* lead to strongly suboptimal coordination outcomes for this second example organization as a matter of principle, the approach of a lax treatment backed by formal and informal rules does at least *allow* to reach better overall outcomes. The ultimate result, however, primarily depends on the effectiveness and efficiency of the motivational instruments. Table 8.3 represents the case of optimization mechanisms like internalization and technical support being successfully applied to the motivational mechanisms and thereby lowering the respective costs to a certain extent.

Altogether, our second idealized example organization should at least abstain from

[9] This would also be the case for any imaginary approach trying to explicitly decide over *any* possible case of hotspot usage and to enforce such a highly differentiated coordination outcome through architectural means. Due to the strong information asymmetries and uncertainties, such an approach would lead to immense efficiency losses, too. Recall, in this context, the considerations from section 3.2.1 and from chapter 6.

[10] These mechanisms include, in particular, technical support and rule-internalization. See sections 7.2.2.1 and 7.2.3.1 for further details.

Table 8.3: Costs of different approaches for WLAN-security – Example 2

	General Ban	Login Automation	Lax Treatment
Coordination Effort	none – low	low	low
Risk of Maladaption	very high	high	low – medium
Fixed Motivation Costs	medium	medium	low
Ultimate Enforcement	none	none	medium – high
Residual Loss	none – low	none – low	medium

the first approach of a general ban. Even if it would lead to virtually nonexistent coordination efforts being necessary and even if motivation costs would be effectively confined to a medium level of initial fixed costs, the specific requirements of the organization would result in an immense risk of maladaption being present because of a highly suboptimal coordination outcome. The coordination process that underlies the general ban would simply not allow to take enough situation-specific information into account and to derive the necessary, strongly diversified results. Login automation, in turn, would lessen the risk of maladaption to a certain extent while leaving all other costs basically unchanged. And finally, a lax treatment would prevent maladaption costs to a great extent but would on the other hand cause considerable motivation costs, including the residual losses arising from nonconformance. The ultimate decision between "login automation" and "lax treatment" should thus depend on the amount of possible benefits that could be derived from an increased number of hotspots being usable and on the organization's ability to actually profit from optimizations of the motivation process.

8.4 Conclusion

As this chapter has shown, the initial problem of organization-internal information security with regard to the access of publicly usable WLAN hotspots outlined in section 1.2 can also be analyzed on the basis of the abstract concepts developed so far. Basically, the (potential) use of WLAN-hotspots requires cooperation among the members of an organization to reach efficient outcomes. Like any other cooperation, this requires the subtasks of *coordination* and *motivation* to be solved. The three intuitive approaches identified in section 1.3 do, on an abstract level, represent different ways for realizing exactly this.

A "general ban" is based upon strong generalization during the coordination process, which generally leads to considerable risks of suboptimal coordination outcomes. "Login automation" allows for a slightly more diversified coordination process and thereby reduces the respective costs resulting from maladaption to a certain extent. In both cases, motivation is realized through architectural means, leading to some ini-

tial fixed motivation costs but raising effectively no costs for any single act of ultimate enforcement and causing no substantial residual losses. A "lax treatment", in turn, uses delegation to minimize the coordination costs of maladaption but requires motivation costs to be borne for ultimate enforcement and also causes a certain amount of expected residual losses as a consequence of the respective motivational instruments' high violability.

These highly generalized statements must always be seen in the light of the specific givens of the organization that is considered. As it was also shown in this chapter, some organizations might not really suffer from the high risk of potential maladaption being present for the first approach of a "general ban" but might on the other hand face vast potential costs for the case of the motivational instrument permitting residual losses at all. Instead of implementing a "lax treatment", such organizations should rather use the approach of a "general ban" or of "login automation". For other kinds of organizations, the negative effect arising from suboptimal coordination might be the pivotal issue because of WLAN-connectivity being an essential prerequisite for value generation while the potential risk arising from opportunistic member behavior might be considered less relevant. Such organizations should rather follow the approach of a "lax treatment" backed by formal and informal rules than implementing "login automation" or even a "general ban".

Regarding the question which of the three intuitive approaches should be used by an organization, we could thus close our whole considerations with a simple statement here, saying something like "It depends." A multitude of factors could be itemized as having to be taken into account, including the importance of mobile connectivity for value generation, the actual risk arising from the possibility of an adverse event actually happening, the organization's ability to profit from mechanisms like rule-internalization and so forth. But such a statement would unquestionably be a poor result for the extensive amount of pages spent so far. Such a statement could already have been given by intuition and without any need for further abstract considerations as those carried out in the preceding chapters. The remarks on the different relations between the level of security and the level of opportunities already outlined in section 1.3.5 would basically have led us to the same result, negating any new insights to be provided by all the theoretical reflections made so far. If we closed here with a statement like "It depends", the whole work would be of at least questionable value.

There is, however, an aspect that was already mentioned but not examined in detail so far: the continuous increase of information asymmetries and uncertainties that characterizes typical scenarios of hotspot usage. This aspect, its influence on the costs of motivation, and its broader implications for the realization of security-related cooperation shall be the subject of the following chapters. As we will see, there *is*, in fact, a deeper insight hidden in the considerations made so far.

Conclusion:
The use of public WLAN hotspots and the different approaches for realizing information security in this context can also be analyzed on the basis of the abstract concepts developed so far. Basically, the use of hotspots represents a further increase of information asymmetries and uncertainties. Organizations can respond to this development in various ways. Those organizations that would suffer from the possibility of opportunistic member behavior more than from a the possibility of maladaption should use the strategy of a general ban or of an automated login. Other organizations, for which the availability of mobile connectivity is essential, should, depending on their ability to use formal and informal rules efficiently, consider to practice a lax treatment instead.

Part III

Solving the Puzzle

Chapter 9

Problem Re-Generalization

Chapter 9
Problem Re-Generalization

[D]id you figure out the head fake?

– Randy Pausch

After ultimately having returned to the initial problem of public WLAN hotspots, let us again give a brief summary over the status reached so far. In part I, we explicated the basic principles being relevant for our model of information security inside organizations. In particular, we developed the two schemes of "computing eras" and of the security "triangle", which restructure the field of information security from two different perspectives. Furthermore, we examined the economic fundamentals of cooperation among different individuals, which can basically be realized in two different ways: through hierarchies or on the basis of market mechanisms. In both of these cases, cooperation causes costs which can in turn be subdivided into coordination costs – the costs arising for the process of determining an optimal state of behavior – and motivation costs, which have to be borne for motivating individuals to actually behave in accordance with the coordination outcome.

Part II, in turn, was devoted to the development of an economic understanding of information security inside organizations. Chapter 5 showed that organization-internal information security requires cooperation among the different members and that this cooperation is primarily realized in a hierarchical manner nowadays. To develop an economic understanding of security-related cooperation inside organizations, we thus had to consider the respective hierarchical cooperation costs in more detail. Consistently with the distinction of coordination and motivation costs from the first part, we then discussed the coordination costs of information security and their development over the different "computing eras" in chapter 6 and identified the different "meta-measures" from the security triangle as motivational instruments with strongly different cost structures in chapter 7. In chapter 8, we finally reconsidered the initial case of public WLAN on the basis of the abstract concepts developed so far. This logical structure is depicted in figure 9.1.

On a highly abstract and generalized level, we can aggregate our findings as follows: In the very beginnings, information security had to be realized with regard to single, isolated systems. Only very few relations had to be considered and each of these relations was of low complexity, ultimately leading to a nearly perfect coordination outcome and thus to a very low level of coordination costs. The coordination outcome

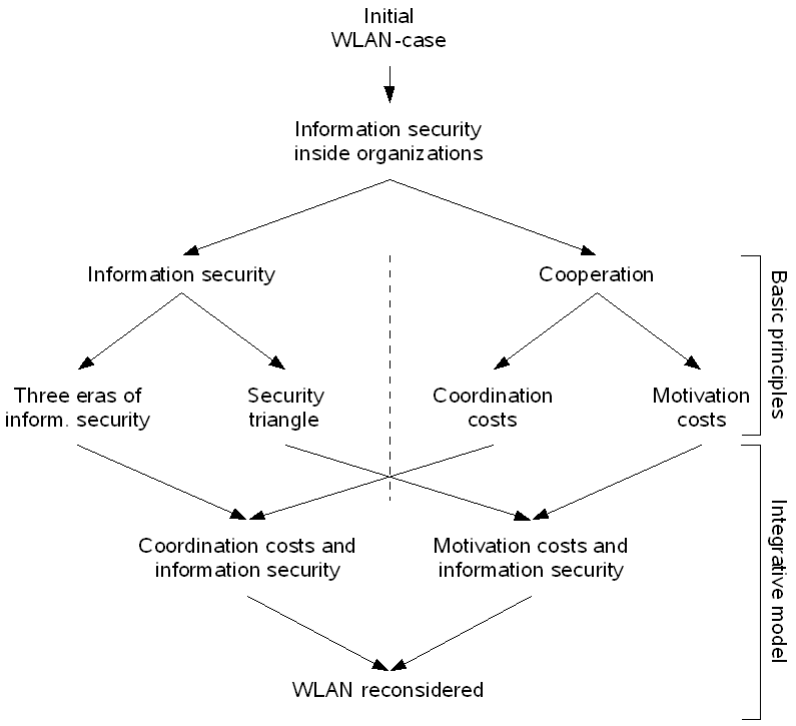


Figure 9.1: Structural overview of status reached so far.

was then enforced by architectural means of motivation, raising a certain amount of initial setup costs but nearly no additional costs of ultimate enforcement.

The introduction of shared, terminal-based mainframe systems then constituted an increased complexity of the coordination process. Due to interrelations between different users or processes being possible, coordination had to be realized in a more differentiated manner. Instead of member-system relations, the coordination process now had to be realized with regard to member-resource relations and different kinds of access (read, write, append, etc.). This increase of relevant relations together with the newly arising possibility of interdependencies and the heightened level of information asymmetries between the centralized coordination instance and the individual member increased coordination costs and had to result in certain amounts of maladaption. Motivation was, however, still realized on the basis of architectural means, even if mechanisms like access control systems were now used instead of the former physical means. The basic structure of motivation costs was thus generally the same.

Even if organizations were basically able to derive profits from the use of shared mainframe systems instead of isolated systems, a certain amount of these profits was thus neutralized by an increase of cooperation costs which, in turn, mainly consisted of increased coordination costs as a result of heightened complexity and asymmetries. Due to these cooperation costs, the real potential of the newly introduced systems could – compared with the isolated systems of the first era – be exploited less com-

prehensively.[1] The remaining value must, however, still have been larger than for isolated systems. Mainframes wouldn't have prevailed instead. These cost relations can be illustrated as done in figure 9.2.

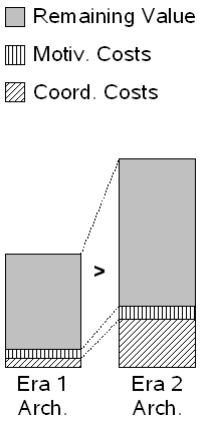


Figure 9.2: Change of cost structure from era 1 to era 2.

The advent and establishment of PCs then led to another change of cost structures. Complexity and asymmetries increased even further and unknown entities now had to be taken into consideration, too (see section 6.3). From the perspective of cooperation costs, this would have induced a significant rise of the coordination part if the established paradigm of architecture-based enforcement of a well-defined coordination outcome were upheld. Either the organization could have tried to reach a highly detailed and well-defined state of aspired behavior and had to accept strong efficiency losses in doing so, or the organization could have based the coordination process upon strong generalization, leading to significant maladaption costs. Both options – as well as the most probable optimized state of medium detailedness – would, as compared to the former era of mainframe-systems, have led to strongly increased coordination costs. Even if the motivation costs for ultimate enforcement were assumed as being essentially constant, this rise of coordination costs would in a multitude of cases have exceeded the increase of value being generally possible through the use of PCs instead of mainframe-systems. Ultimately, this would have led to *decreased* remaining profits and thus militated against actually using PCs in an organizational context. Figure 9.3 illustrates this relation.

Within this understanding, affected organizations thus had two options for dealing with the possibility of using PCs: They could either have gone without them and kept using mainframe-systems, or they could have searched for alternative approaches for realizing secure member behavior that did not show the deficit of substantive coordination costs having to be borne and that left larger remaining profits than the

[1] Actually, the *whole* potential could also not be exploited during the first era of isolated systems because of some amounts of coordination and motivation costs being present, too. But these losses were less significant than for the much more complex world of shared mainframe systems.

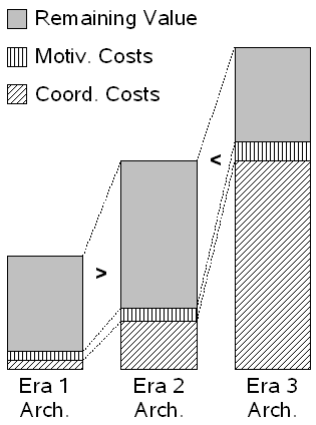


Figure 9.3: Change of cost structures from era 1 to era 3.

use of mainframe-systems did. Actually profiting from the increased possibilities of PCs would only have been possible if such an alternative approach could be found.

The additional use of formal and informal rules instead of merely architectural means represents such an alternative approach: By delegating a certain amount of ultimate *coordination* decisions to individual members (see section 7.4), organizations could counteract the negative effect of increased complexity and information asymmetries to a certain extent, thereby alleviating coordination costs significantly. This delegation did, however, require additional *motivation* costs to be borne. On the one hand, the ultimate enforcement of formal rules is more expensive for the single case than enforcement through architectural means and on the other hand, a certain amount of rule-breaking behavior has to be expected because of the high violability of these meta-measures. Furthermore, the motivation process must not only refer to a certain coordination outcome to be enforced. Individual members also have to be motivated to make coordination decisions that coincide with the collective interest for delegation to work in the interest of the organization in the above-mentioned sense.

If, however, this alternative, mixed approach actually led to an overall structure of cooperation costs that left a larger amount of remaining value for the organization than the use of mainframe-based systems with architecture-based enforcement – a visualization is given in figure 9.4 – then the use of PC-based sytems instead of mainframes would in fact have been economically reasonable for the organization. Even if being anything but a proof, this does at least lead us to a plausible explanation for the shift from a mainly architecture-based approach to the mixed strategy also resting upon formal and informal rules that happened together with the shift from the second to the third era: Only through a shift of meta-measures being used and through the mechanism of delegation being made possible by this shift, organizations could counteract the increase of complexity and asymmetries and actually derive overall profits from the use of PCs instead of mainframe-based environments.

Now, reconsider the specific characteristics of the WLAN-case that were already

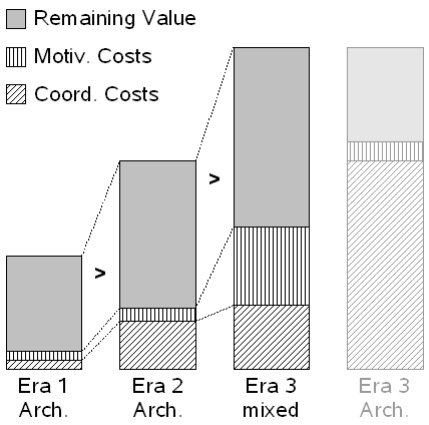


Figure 9.4: Relative cost structure of era 3 with "mixed" approach.

outlined in section 8.1: A further increase of complexity, uncertainty and information asymmetries, a breaking of network boundaries and, in particular, a strongly increased relevance of situation-specific aspects. Within the abstract and highly generalized model outlined so far, these specifics would give reason to use formal and informal rules instead of architectural means in order to minimize overall cooperation costs. But during the discussion of this approach in section 8.2.3, we also identified another aspect that was not examined in detail so far: As typical WLAN-settings involve a high level of asymmetries between the individual member and the centralized instance as well as between the different members, motivation costs for formal and informal rules are – compared with classical office settings – increased significantly. Both possible options, realizing the same probability of nonconforming behavior to be detected on the one and lowering the probability on the other hand would ultimately result in higher motivation costs having to be borne by the organization. Like coordination costs increased significantly with the change from mainframe- to PC-based environments, the increasing use of WLAN-hotspots in mobile settings thus increases *motivation* costs.

9.1 The General Trend of Decentralization

If, however, the WLAN-case is not interpreted as a specific, isolated case but rather as *just one* manifestation of a much broader development, then there is a deeper insight hidden in the considerations made so far, leading to potentially far-reaching implications for the field of organization-internal information security. The indications for such a broader development to be currently getting under way are manifold: In section 3.4, for instance, we already addressed the impact of technological change to organizational structures and mentioned different studies that suggest a size-*decreasing*, decentralizing effect of information technology resulting from lowered transaction costs

to prevail over the size-*increasing* effect of hierarchical cooperation being realizable more efficiently, too.

Other observations point into the same direction of a decentralization and de-hierarchization of economic activities taking place. One of the most obvious respective phenomena is the growing importance of mobility for the conduct of knowledge work. Instead of residing within well-defined and well-known office environments, knowledge-workers are increasingly acting within mobile and ever-changing environments. This mobility cannot be narrowed down to the use of WLAN hotspots but rather includes a wide variety of scenarios, ranging from public transport being used as temporary workplace over occasional homework days and off-site conferences to work-related meetings being held in bars or cafés. The underlying development has been widely discussed in the scientific world under such keywords as *"mobile (tele-)work"*[2], *"nomadic computing"*[3], *"mEnterprise"*[4] or *"distributed work"*.[5]

The benefits that can be derived from such mobile, locally dispersed work-practices can be manifold and shall not be discussed in detail here.[6] But on the other hand, mobile scenarios typically require organization-internal information to be accessed, edited or created outside of the formerly well-established physical boundaries of the organization. Even without any network connection being present, internal information is taken out physically to conduct work in a park, on a train or in a café, thereby effectively blurring the distinction between the "inside" and the "outside" of organizations.

The increasing use of outsourcing is another practice resulting from technological progress that changes established organizational structures. Instead of conducting all or most necessary activities within well-defined boundaries through basically hierarchical arrangements, organizations nowadays make extensive use of external contracting. This refers to specialized outsourcing providers as well as to the increasing importance of freelancers, temporary members or consultants.[7] The status of the respective individuals is less clear than it had been for "classical" organizations: A freelancer working for the organization in a specific project, for example, can neither be considered as complete non-member of the organization nor is he a member in the classical, established sense. An "external" consultant, in turn, might primarily be seen as a member of the consulting organization but on the other hand, she might be strongly involved into organization-"internal" cooperative processes and play the temporary role of a (part-time) member of the client organization. And with an increasing amount of formerly internal functions being relocated to outsourcing providers, members of the

[2] See, for example, Perry, O'Hara, Sellen, Brown, and Harper (2001), Brodt and Verburg (2007) or the collection by Andriessen and Vartiainen (2005).

[3] See, for instance, Kleinrock (1995, 1997, 2001) or Lyytinen and Yoo (2002).

[4] See Gould, Jackson, Schyndel, and O'Donnell (2006).

[5] See Kallinikos (2007, pp. 91f), Yates, Orlikowski, and Woerner (2003) or, as another collection, Hinds and Kiesler (2002).

[6] To mention just a few, individual members can better react to situation-specific requirements, the working environment might be more inspiring or less disturbing, the relation with customers or partners might be improved, the organization could profit from a lessened number of fixed office workplaces having to be provided, traveling time could be used more productively, etc.

[7] See, as representatives for many others, Drucker (1999, p. 15) and Malone (2004, pp. 31ff, 74ff).

"external" contracting organization and their behavior become increasingly relevant for the "internal" process of cooperation, too.[8]

From an abstract perspective, these developments blur the organizational boundaries even further: Organization-"internal" information is increasingly administrated and processed by "external" entities, "external" individuals are more and more involved in "internal" processes and membership is of temporary nature more often than it had been before. Information technology and its progress do again play a significant role as necessary enabler for such developments.[9]

Finally, current trends in information technology also suggest advanced decentralization, de-hierarchization and the blurring of organizational boundaries to be expected. The paradigm of *"Software-as-a-Service"* (SaaS), for instance, can be seen as a specific case of outsourcing and is increasingly promoted by vendors such as Google, SAP or, as one of the most interesting examples, Salesforce.com. In any of these cases, large amounts of information that was formerly considered as *strictly* internal are now stored, processed and analyzed within the infrastructure of the respective "external" vendor while the amount of IT-based operations being conducted "internally" does, in turn, decrease.[10] The distinction between "internal" and "external" activities thus gets even more ambiguous. Closely coupled with the concept of SaaS is that of service-oriented architectures (SOA). These are primarily promoted as a possibility to enhance flexibility and to foster the reuse of software components. Instead of building monolithic applications, the underlying idea is to assemble large parts of software solutions from existing services. These services can, in turn, be operated within an organization as well as they can be obtained from outside entities.[11] For the latter case of "external" services being used, the effect of such service-oriented architectures is again blurring the line between "internal" and "external".[12]

Without going more into detail, all these and many other current developments represent more decentralized and less hierarchical ways of conducting economic activities. The boundaries between the inside and the outside of organizations are slowly disappearing in a multitude of ways[13], new forms of work increasingly render the binary distinction of "members" from "non-members" inappropriate and IT-infrastructures are incrementally relocated to what would formerly have been called "the outside". Even the term of an "organization" itself becomes vague to a certain extent. Together with the considerations from section 3.4, this suggests that more and more organiza-

[8] Note that this does not only apply to the various scenarios of formerly "internal" functions and information being relocated to external parties but also to those cases of outsourcing where "external" agents access "internal" resources.

[9] See, for instance, Malone (2004, p. 31), Brynjolfsson et al. (1994) or Kallinikos (2007, pp. 92 f, 101 ff).

[10] In this context, see also Carr (2005), predicting *"the end of corporate computing"*.

[11] For a short introduction, see, for example, Papazoglou and Georgakopoulos (2003) or Huhns and Singh (2005).

[12] Bieberstein, Bose, Walker, and Lynch (2005, pp. 697 ff), however, also suggest that even for the case of a SOA solely being built upon internally provided services, the full potential can only be exploited with accompanying changes of organizational structure.

[13] For additional comments on the same abstract issue, see also Brynjolfsson and McAfee (2007, p. 54) or Chesbrough (2003).

tions will be transformed from well-defined and clearly delimited entities into *"loosely coupled organic networks"* (Dhillon and Backhouse 2000, p. 125). Traditional ways of cooperation – that is, hierarchies – will thus increasingly be replaced by alternative arrangements.

Of course, these developments do not necessarily have to apply to any area of economic conduct in the same manner and of course, they are everything but already representing the predominating status quo. But on a generalized basis, there are strong indications for them to be currently gaining momentum and to change established practices of economic cooperation among different individuals.

What, then, does this implicate for the field of organization-"internal" information security? This question shall be discussed in the following.

Conclusion:
Ongoing technological develoments lead to further changes for the prevailing ways of conducting economic cooperation. The increasing importance of mobility, outsourcing and other novel work forms and even primarily technological trends like *"Software as a Service"* (SaaS) and *"service-oriented architectures"* (SOA) tendentially lead to blurring organizational boundaries, non-binary modes of "membership" and to cooperation arrangements that strongly differ from the established way of hierarchical cooperation. Information security will have to take these changes into account.

9.2 Decentralization and Information Security

In a first step of analyzing the above-mentioned developments within our model developed so far, we can identify a fourth "era" of computer usage as currently gaining momentum. This era is primarily shaped by a paradigm of "interwoven"[14] computing entities: Central facilities of different organizations are interconnected with each other, mobile devices are used within the organization-"internal" network as well as they are connected to "external" entities like WLAN hotspots, home networks or even the "internal" networks of other organizations. Individual consultants can be "members" of different organizations at the same time or in quick succession and do of course have a need to access "internal" information of any of these parties. And finally, all the different forms of outsourcing or "external sourcing", including SaaS and service-oriented approaches, also lead to much more interwoven structures and less clear organizational boundaries than it had been the case before. Figure 9.5 shall illustrate this newly arising computing paradigm.

Of course, these developments have not remained unrecognized by information security scholars and professionals. Above all, the increasing mobility of an organization's members has been subject to a wide variety of security-related considerations. In the beginnings, the challenge of realizing information security in this context was primarily

[14] Others, like Malone (2004) or Castells (2000), would presumably prefer the term "networked" here. This would, however, not make the distinction from the third era clear enough. The term "interwoven" seems to be more appropriate in this respect.

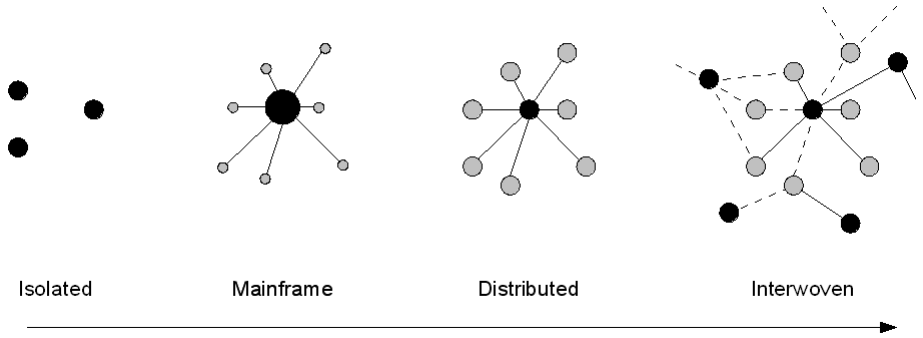


Figure 9.5: The fourth era of "interwoven" computing.

understood as necessitating new technical methods for issues such as authentication, data encryption, storage management, etc.[15] The same, primarily technical perspective also prevailed during the first years of the third millennium. Eustice, Kleinrock, Markstrum, Popek, Ramakrishna, and Reiher (2003), for example, propose a mechanism based on *"quarantine, examination and decontamination"* for approaching the problem of mobile users traveling between different wireless environments.

Recently, however, the mentioned developments have also been discussed from a more abstract perspective. Booker (2006, p. 15, table 1), for instance, recognizes – in consistence with the above perception of blurring organizational boundaries – the increasing mobility as *"erod[ing] the traditional perimeter security model"* that many organizations pursue nowadays. Palmer (2005), a representative of the Jericho Forum[16], does not only refer to mobility but also considers some of the above-mentioned further phenomena of decentralization and also recognizes a strong trend toward *"de-perimeterisation"* that has to find expression in changed security practices.

Even if this view has also been taken by a multitude of further authors[17] and the progress achieved by all these efforts notwithstanding, the actual effects for practical information security as well as the impact on information security research are still unsatisfactory. The problem of information security with regard to mobile users and the upcoming new organizational forms is still unsolved to a large extent. There is no established abstract concept that explicitly addresses the specific security issues arising from the above-mentioned developments and that could serve as basis for well-founded scientific as well as practical considerations. Even if the abstract problems are widely recognized, solutions or even solution approaches are still rare – if not nonexistent.

[15] See, for instance, Forman and Zahorjan (1994, pp. 41 f, 44 f), Bharghavan and Ramamoorthy (1995) or Mavridis and Pangalos (1997).

[16] The Jericho Forum (see `http://www.opengroup.org/jericho/`) is a panel within the Open Group concentrating on the impact of de-perimeterization for the field of information security. See also the various publications at `http://www.opengroup.org/jericho/publications.htm` [24.11.2008] and, in particular, Lacey (2005, p. 187).

[17] See, amongst many others, Bouchard (2004, p. 5), Knight (2004) or Kriens (2005, p. 3). See also Geer (2008, p.32), stating that *"the term 'corporate perimeter' largely becomes an oxymoron."*

The specifics of the fourth era of interwoven computing shall thus be analyzed by means of our economically inspired model developed so far. Throughout this analysis, we will again confine ourself to the consideration of abstract, generalized implications arising from the progression toward the fourth wave of *"interwoven"* systems. Consistently with our abstract model, we will distinguish between the implications for the *coordination* process (section 9.2.1) and those affecting the *motivation* process (section 9.2.2). Section 9.3 will then elaborate on the implications for information security that arise from these abstract considerations. The subsequent chapters will ultimately present and discuss different possible approaches for overcoming the newly arising challenges.

Conclusion:
The ongoing trend toward more mobility, decentralization and de-hierarchization has already been recognized as important subject for future directions in the field of information security. Established practices of perimeter-focused security increasingly turn out to be inadequate. Well-founded alternative solutions, solution approaches or abstract concepts are, however, virtually nonexistent.

9.2.1 Decentralization and Hierarchical Coordination of Information Security

With regard to the hierarchical process of coordination, the implications of the switch from the third to the fourth era are quite obvious. Remember that coordination refers to the determination of a state of desired security-related member behavior that maximizes the organization's overall benefit. As outlined in section 3.2.1, this process produces costs which, in turn, strongly depend on the number of relations having to be considered, the need for repeated reconsiderations, the complexity of each singular decision and the level of uncertainties and information asymmetries existing between the centralized coordination instance and the individual member (see sections 6.1 to 6.3). The switch from the third era of distributed computing to the fourth era of interwoven systems affects all these factors.

The number of relevant relations, for example, is increased because of the multitude of previously irrelevant "external" entities having to be considered, too. Members of an organization connect to unknown "external" networks and access "external" information resources, "externally" obtained services might have to be considered with relation to "internal" information and "external" individuals perform access to "internal" entities. The increasing discontinuity of "membership" heightens the number of relations even further: As any change of membership status comes along with new relations having to be considered or existing relations having to be discarded, the overall number of relevant relations does again increase. On the other hand, the number of formerly "internal" relations might, in the course of relocation to external parties, be *decreased* to a certain extent. An additional relation between a member and an externally hosted, SaaS-based CRM-system, for instance, will usually come along with the elimination of the respective relation between the member and the formerly used in-

ternal system. The *overall* number of relevant relations, however, will hardly decrease significantly. Instead, it will either stay on a level comparable to the third era or it will – and this is more likely – increase to a certain extent.

Even more relevant than the increase in the number of relations having to be considered is the strongly amplified need for repeatedly reconsidering each of these relations. It was, for example, already outlined above that it could very well make a difference for the determination of the aspired behavior whether a certain member currently resides in a private hotel room or in the lobby of the same hotel. Due to this context-relevance, it is not enough to determine whether a certain member "should" perform a certain kind of access to a specific resource or not. Instead, we will have to ask if he should do so within a specific situation or not. A demonstrative example might be the reading access to some locally stored business numbers for the purpose of preparing a presentation. While we only had to determine whether a specific member "should" or "should not" read these numbers during the third era, we now and in the future have to ask if he "should" access them in a hotel room where the risk of shoulder-surfing is comparably low and whether he "should" do so in other contexts with a higher risk of shoulder-surfing – while sitting in the hotel lobby or on a train, for instance. For any significant change of context, the relation thus has to be reconsidered to reach an optimal coordination outcome.

The complexity of each of these relations and the respective decisions having to be made will presumably increase, too. While during the first era of isolated systems it only had to be decided whether a certain member of the organization is an operator of the considered system or not, the shared systems of the second era required the possibility of interdependencies between different users, processes and resources to be explicitly taken into account during the coordination process. The third era then led to an increased complexity of each decision because of the heightened user flexibility and the possibility of even unknown devices and information resources being relevant for the coordination process, too. The fourth era, however, heightens this complexity even further. For the case of mobile users, for example, the respective current context (see above) had – besides all the issues already known from the third era – to be taken into account in a multitude of ways for the (hypothetical) aim of realizing "perfect" coordination. Deciding whether an external consultant "should" access a certain information resource might also depend on the question for which other clients the same consultant also works. And with service-oriented architectures resting upon externally obtained and complexly intertwined services, a wide variety of formerly irrelevant factors – the actual locus of service provision or some specific knowledge about how the service is realized internally, for example – will have to be taken into account as well. All these factors will render each decision more complex than before.

Finally, the level of uncertainty and information asymmetries under which all these highly complex relations have to be evaluated also increases. While the former eras were at least characterized by well-defined and well-known environmental conditions for member behavior – the given office environment, for example – the characteristics of the fourth era lead to large parts of the centralized coordination process having to be realized under conditions of extensive lacks of knowledge. Not only will any singular relation become more complex to evaluate because of the respective context

being highly relevant, but also will the coordinating instance hardly possess the necessary information about the respective context-specific givens and requirements. The coordination instance will neither be able to predict any possible context that could become relevant for decision-making[18], nor could it make well-founded assumptions about the internal workings of externally obtained services, the working environment of outsourcing providers or freelancers, or the level of security provided by an SaaS-application. Besides the increased number of relations having to be considered, the increased need for reconsideration and the higher complexity, any single relation will thus have to be evaluated on the basis of a tendentially higher level of uncertainty and information asymmetries.

With the switch from the third era of "decentralized" to the fourth era of "interwoven" systems, all primary factors determining efficiency losses will thus presumably develop in a direction that *increases* the costs of hierarchical coordination even further. The rising number of relations having to be considered together with the increased complexity leads to the capacity limits of the centralized instance being reached earlier and the heightened level of uncertainty and information asymmetries leads to decisions being increasingly suboptimal.[19] For both reasons, increasing efficiency losses can be expected on a generalized basis (see table 9.1[20]).

Table 9.1: Basic model of efficiency losses of hierarchical coordination – Comparison of eras 2 to 4

	Era 2	Era 3	Era 4
Number of necessary decisions	medium – high	high – very high	↗
Need for repeated reconsideration	medium	medium	↗
Complexity of each decision	medium – high	high – very high	↗
Level of uncertainty / lack of information	medium	very high	↗
Overall efficiency loss	*medium – high*	*high – very high*	↗

[18] In this context, see also Dhillon and Backhouse (2000, p. 128), noting that established organization-internal rules *"apply in foreseen and predictable circumstances and cannot be invoked in new and dynamic situations"* and thereby indirectly illuminating the underlying coordination problem.

[19] See also Denning and Hayes-Roth (2006), referring to *"Hyper-Networks"* and challenging the approach of trying to manage them centrally: As there is *"too much data to aggregate"* (p. 23) in order to identify some kind of optimal state and for various other reasons, hierarchical decision-making leads, according to Denning and Hayes-Roth, to strongly suboptimal outcomes and should thus be replaced by alternative models of decentralized decision-making.

[20] The table lists the basic estimations without any optimization being present (see tables 6.2 on page 122 and 6.4 on page 133). Optimization is considered below. The first era is omitted for layout reasons alone. If relevant, please refer to table 6.1 on page 118 instead.

Like it had been the case for the second and the third era, these losses can be decreased to a certain extent by means of optimization. Generalization was already mentioned in sections 6.2.1 and 6.3.1 and could also be applied here. A centralized instance could, for instance, decide that no external services or no services being hosted outside the organization's home country should be used for composing internal service-based applications *in general* because of availability considerations or general doubts about the legal enforceability of contracts. It could make the generalized decision that mobile devices should *never* be used outside of private locations like hotel rooms because of the omnipresent risk of shoulder-surfing or it could come to the generalized conclusion that *no* "external" consultant or freelancer should access *any* information related to R & D projects. Such generalized decisions would again reduce the efficiency losses arising from capacity limits being reached but would on the other hand also avoid behaviors that would be beneficial for the organization. Generalization might – if used cautiously – thus again limit the efficiency losses to a certain extent, but if exaggerated, it would itself lead to strongly suboptimal coordination outcomes.[21]

The same analogy to the previous eras applies to the optimization approach of delegation. The centralized instance could counteract the risk of capacity-limits being reached – which arises from the vast number of relations having to be considered and the increase of all other factors mentioned above – by delegating coordination decisions to lower layers such as department-specific security officers or, more related to the particular characteristics of the fourth era, to specialized security officers for mobility, "external" services, or cooperation projects, for example. As already outlined above[22], this would not only allow to overcome the problem of capacity limits but also to alleviate the strong uncertainties and information asymmetries to a certain extent. A separate security-officer being responsible for all members of the sales department would, for instance, be more able to weight the benefits of sales-related mobile scenarios against the respective risks than a central instance could. On the other hand, such delegation would in turn give rise to the possibility of interdependencies with other departments not being addressed properly and thus cause another kind of costs itself.

A third approach for optimizing over coordination costs was also mentioned in section 6.3.1: technical support. Enterprise systems management suites, sophisticated reporting solutions and comparable instruments are possible ways for heightening the effective capacity limits and for reducing control loss. Other technical solutions like network scanners were mentioned as being able to lower the level of uncertainty and information asymmetries the coordination instance has to cope with. Of course, such solutions can to some degree also be applied to the fourth era. Enterprise suites might allow to model "external" relations and interdependencies in detail and future solutions may even "comprehend" the ever-changing context to a certain extent, thereby limiting uncertainty and information asymmetries. On the other hand, technical support also has its limits and it is unlikely that, for example, technical support will soon allow an immediate re-coordination for the case of a new and formerly unknown context

[21] Recall, in this context, the relation between those costs arising from efficiency-losses and those resulting from over-generalization shown in figure 6.3 on page 125.

[22] See, for example, section 6.3.1.

having to be included in the respective coordination. Even though technical support might thus reduce the newly arising coordination costs, this optimization will, like the above-mentioned mechanisms of generalization and delegation, never allow to nullify the changed givens that characterize the fourth era completely.

Generally speaking, all the potential drawbacks of optimization already outlined in section 6.3.1 do thus apply here, too. Optimization might reduce efficiency losses but on the other hand causes some losses itself. If being applied properly, optimization might allow a certain reduction of overall losses but even the best optimization will not allow to reduce hierarchical coordination costs to the level that prevailed during the third era. This would require *all* effects of decentralization and "interwovenness" – the increased number of relations, the permanent need for reconsideration and the heightened level of uncertainty and information asymmetries – to be counterbalanced in their entirety by optimization measures without any additional resulting drawbacks. This is at least unlikely to happen and consequently, overall coordination losses will, compared to the third era, increase even further.

As a final aspect of coordination costs, we have to consider *bureaucracy costs*. As already outlined in sections 3.2.1 and 6.4, these costs arise from decision-makers typically overestimating their own abilities and furthermore pursuing own goals. Managers striving after a higher social status or higher payments because of more responsibilities are just two examples for such a *"strategic propensity to manage"* (Williamson 1985, p. 149). With regard to the hierarchical coordination process of information security, this can lead to less delegation of security-related coordination being used than would be reasonable, to coordination decisions being made on the basis of strategic deliberations or individual preferences instead of conscious considerations, or to a multitude of other adverse effects for the hierarchical process of coordinating security-related member behavior. Other factors like *"forgiveness"* and *"politicization"* (Williamson 1985, pp. 150 ff) point into the same direction.

Under the conditions of strong decentralization and interwovennes that can be expected for the fourth era, these bureaucracy costs are likely to increase even further. As explained at the end of section 6.4, factors like information asymmetries, uncertainties and complexity positively affect the resulting bureaucracy costs because the coordination instance has higher incentives to safeguard the own position against unforeseen events, for example. As all these factors can be expected to increase with the imminent shift from the third to the fourth era of computing (see above), we can thus expect an even higher level of bureaucracy costs having to be borne by the organization.

Taking all these considerations into account, we can thus expect *all* factors of hierarchical coordination costs for determining the "optimal" state of security-related member behavior to increase from the third to the fourth era. Even if optimization can again limit the cost increases to a certain extent, it will only hardly allow to counterbalance them entirely. Bureaucracy costs also have to be taken into account and will presumably increase from the third to the fourth era, too. This leads us to a cost structure for the hierarchical coordination of security related member behavior in the fourth era that is aggregated in table 9.2: Starting from the current world of the third era, we can – based on our abstract considerations – expect all factors of hierarchical

coordination costs to increase even further.

Table 9.2: Overall losses of optimized hierarchical coordination – Comparison of eras 2 to 4

	Era 2	Era 3	Era 4
Number of necessary decisions	medium	medium – high	↗
Need for repeated reconsideration	medium	medium	↗
Complexity of each decision	medium	high	↗
Level of uncertainty / lack of information	low – medium	high	↗
Additional drawbacks due to optimization	low – medium	high	↗
Overall efficiency loss	*(low –) medium*	*(medium –) high*	↗
Bureaucracy costs	low	medium	↗
Overall costs of hierarchical coord.	*medium*	*high*	↗

This would, in turn, argue for a further delegation of ultimate decisions to the individual "members" in the sense of section 7.4: Instead of trying to coordinate behavior in a centralized manner directly, the ultimate decisions could be transferred to the individual who is much more able to know the situation-specific givens and requirements and thus does not face all the drawbacks outlined above. This individual, however, had to be motivated to make decisions that comply with the collective interest. Formal and informal rules have been suggested to serve exactly this goal and consequently, it would at first sight seem reasonable to increasingly realize information security on the basis of these meta-measures in the fourth era.

Nevertheless, this would give rise to a conflict that had not been present during the previous eras and that will almost certainly require an additional major shift of information security practices to happen when the developments toward the fourth era gain further momentum. As we will see in the following section, there are at least strong reasons to believe that the established, mixed approach from the third era will not be feasible within the highly decentralized and interwoven world of the fourth era anymore.

Conclusion:
Based on our economically inspired model developed so far, we can identify a further increase of all factors determining the height of coordination costs to be likely to come along with the shift from the third to the fourth era of interwoven computing. Even if optimization might allow to minimize cost increases to a certain extent, there will still be a remaining increase of losses, leading to more suboptimal outcomes of the hierarchical process of identifying the aspired state of security-related member behavior than those already known from the third era. At first sight, this argues for a further delegation of ultimate coordination decisions to individual members.

9.2.2 Decentralization and Hierarchical Motivation of Information Security

As outlined above, the various developments constituting a new, fourth wave of interwoven computing basically lead to an increase of all factors determining the costs of coordinating security-related member behavior in a hierarchical manner. It seems therefore plausible to increasingly realize information security on the basis of formal and informal rules in order to delegate ultimate decisions to individual members while at the same time motivating members to make decisions that reflect the collective interest of the organization as a whole.

On the other hand, this motivation also causes costs itself. As section 7.2 has shown, the different meta-measures – architectural means, formal rules and informal rules – feature specific characteristics in this respect which furthermore depend on factors such as the level of information asymmetries being present in a given situation.[23] To derive founded statements about the whole costs of hierarchical cooperation with regard to security-related member behavior that have to be expected for the fourth era, we thus have to examine motivation costs in some detail, too. Additionally, and as already outlined above, the specific characteristics of the fourth era give strong reasons to believe that delegation of ultimate decisions to individual members will play a highly important role for minimizing hierarchical coordination costs. This delegation is currently realized through the use of formal and informal rules and we shall thus analyze how the hierarchical motivation costs of these instruments will presumably develop together with the shift from the third to the fourth era. Architectural means will be omitted for the moment.

In section 7.2, motivation costs were split into *fixed costs* arising for the initial creation of a motivational measure, *marginal costs* having to be borne for aligning an existing measure to changed coordination outcomes, *enforcement costs* for exerting influence on a single case basis and *residual losses* having to be expected because of members – the existence of the motivational instrument notwithstanding – not behaving in conformance with the coordination outcome. Based on this classification, formal rules have been characterized by a low level of fixed costs, a medium level of marginal costs, high costs of ultimate enforcement and medium residual losses

[23] For the role of information asymmetries, see section 7.3.

having to be expected. Informal rules, in turn, show medium fixed costs, a medium to high level of marginal costs, low costs of ultimate enforcement[24] and a high level of estimated residual losses.

For both, formal as well as informal rules, information asymmetries do furthermore play a significant role. As outlined in section 7.3, the individually perceived costs attached to rule-breaking behavior is for both cases mainly determined by the probability of nonconformance being detected and sanctioned and by the severity of the respective sanctions. Larger information asymmetries, in turn, result in a lower probability of rule-breaking behavior being discovered and thus in a lower effect of the motivational instrument, ultimately leading to higher residual losses having to be accepted.

The difference between formal and informal rules consists in the kind of asymmetries being relevant. For formal rules being enforced by a centralized instance, the relevant factor is the asymmetries between the centralized instance and the individual member. The higher these information asymmetries are, the lower is the probability of the member's rule-breaking behavior being detected and sanctioned and the higher will the expected residual losses be. For informal rules that are enforced through spontaneous and informal control, on the other hand, the decisive factor is the level of information asymmetries among the different individual members. The more the individual members are isolated from each other, the lower is the probability that an individual member's rule-breaking behavior is detected through spontaneous control by the other ones and consequently, the higher are the residual losses that have to be expected as result of rule-breaking behavior actually being present. The overall motivation costs of formal and informal rules are thus strongly determined by different kinds of information asymmetries, but in both cases, information asymmetries affect the effectiveness of the respective meta-measure substantially.

Now recall the specifics of the fourth era that were outlined in section 9.1 above: Mobility and local distribution of work, tendentially more decentralized organizational structures, ever-changing and less clear "membership" statuses and the fact of organizational boundaries blurring in a multitude of ways were mentioned as characterizing the era of interwoven computing. Under such conditions, information asymmetries are hardly about to diminish compared to the third era. Instead, both kinds of asymmetries – those between a centralized instance and the individual member as well as those among the different members – are far more likely to *increase*.

First, consider the asymmetry between a central instance and the individual member that is primarily relevant for the enforcement of *formal rules*. Assume that there is such a formal rule prescribing not to use external storage media except for those cases where its use is unquestionably in the interest of the organization. In classical office settings, the non-compliance with such a rule would be detected by the centralized instance

[24] As already noted in section 7.2.3, informal rules are usually enforced by the different members themselves through mechanisms of spontaneous, informal control. They do thus not raise substantial enforcement costs having to be borne by the centralized instance. The costs being shouldered by the individual members for realizing spontaneous control could, nonetheless, also be considered as costs from the viewpoint of the "organization" because members do not carry out productive work while exerting informal control. As we will see in the following, this would, however, make no difference for the general point that will be made herein.

with a certain probability. This probability unquestionably decreases significantly with an individual member residing in a hotel room far away from, say, the CISO of the organization, leading to a higher level of rule-breaking behavior having to be expected. This mechanism was already outlined for the use of WLAN-hotspots in section 8.2.3 and does, of course, also apply to other kinds of security-related member behavior. Either the costs that have to be borne for realizing the same probability of detection increase significantly or the residual losses do so. In any case, the motivation costs of formal rules can be expected to increase strongly because of the heightened information asymmetries introduced with the change from the third to the fourth era.

The other characteristics of the fourth era influence the motivation costs arising from formal rules in a similar manner. The blurring of organizational boundaries and the less clear statuses of "membership", for instance, will presumably lead to situations where the applicability of a given formal rule is uncertain. It might, for example, be unclear whether an external consultant or the member of an outsourcing provider has to obey the rules of the client organization or if he only has to follow those rules given by his "home" organization. Besides this general possibility of ambiguity, rule enforcement might turn out to be problematic in matters of detecting rule-breaking behavior as well as with regard to the exertion of sanctions. Again, this will in all likelihood increase the costs of ultimate enforcement and the residual losses arising from formal rules in the fourth era compared to the less complex relations that prevailed during the third era.

Fixed and marginal costs for initially creating a formal rule and for aligning it to changed outcomes of the (limited) coordination process, in turn, are not significantly affected by the shift from the third to the fourth era. Setting or reformulating a law-like formal rule will still be realizable at the same costs already known from the third era.

Within our distinction of fixed and marginal costs, costs of ultimate enforcement and residual losses, fixed and marginal costs of formal rules can therefore be expected to remain at a similar level while the costs of ultimate enforcement and the residual losses are about to increase significantly (see table 9.3). Generally speaking, the switch from the third to the fourth era will thus in all likelihood increase the severity of the principal-agent relation already known from the third era and will thus lead to a substantial increase of overall motivation costs related to formal rules.[25]

The motivation costs of *informal rules* are likely to develop in a comparable manner. As outlined in section 7.3, the effectiveness of these norms-like motivational means is strongly influenced by the level of information asymmetries existing among the different members. And with the development from the third to the fourth era, these asymmetries are likely to increase, too. When, for instance, different members do not work together in a common office or building anymore but are rather locally dispersed

[25] Notably, this would also be the case if formal rules were not used as a means of delegation but rather in a highly strict manner, prescribing aspired behavior in a detailed and explicit fashion. But if the highly detailed state of aspired behavior required for such formal rules could be realized as outcome of the coordination process, it could also be enforced by architectural means without the substantial costs of ultimate enforcement and the significant residual losses. Recall, regarding this aspect, the discussion in section 7.4.

Table 9.3: Motivation costs of law-like formal rules – Changes arising from the switch from era 3 to era 4.

Meta-Measure	Fixed Costs	Marginal Costs	Enforcement Costs (single case)	Residual Loss
Formal Rules	→	→	↗	↗

from each other, this implies an increase of information asymmetries between them. Individual behavior that violates an informal rule is unlikely to be detected at all and consequently, as already outlined for formal rules above, residual losses are likely to rise. An informal rule condemning the access of dubious websites, for example, has a completely different effectiveness in an isolated hotel room than in a traditional office environment. The same mechanism also applies to other informal rules regarding the use of external storage devices or even more diffuse ones prescribing "careful behavior" in general: When there is no social control, the effectiveness of informal rules is strongly limited, ultimately leading to a higher level of residual losses having to be expected.

Besides local dispersion of individual members, other aspects characterizing the fourth era point into the same direction of norms-like motivation becoming significantly less effective. Take, for instance, the trend toward more decentralized and ever-changing organizational structures: Within such an environment, the development of some specific informal rule is – even with hierarchical means like those outlined in section 7.2.3 (influencing change-agents, providing detailed information to opinion leaders, etc.) being used – less likely than it has been during the third era. The increasingly important role of part-time and temporary "members" and of "external" consultants will at least lessen the likelihood of norms-like informal rules to emerge and stabilize because of less stable social settings and fewer repeated interactions being realized among the members of these settings. Norms-like informal rules are thus also less likely to emerge at all within the environments that characterize the fourth era of interwoven systems and "organizations", leading to increased fixed and marginal costs having to be borne for establishing new and changing existing informal rules. This inability of organizations to establish or change a certain informal rule – or rather the immense costs that had to be borne for doing so – might also lead to residual losses increasing even further.

The same principle also applies to the problem of ambiguous applicability of a given informal rule already mentioned for formal rules above. Due to their ever-changing "membership" statuses, individuals will increasingly belong to different social groups with different sets of informal rules at the same time, presumably leading to uncertainties or conflicts regarding the question which informal rule to follow in a given situation. Should, for example, an "external" consultant follow the informal behavioral rules of his "home organization" or rather those of the client, which she might additionally not even be able to know because of the temporary nature of her

consulting activities? Again, the specific characteristics of the fourth era will in all likelihood lead to higher residual losses of hierarchical motivation based on informal rules than it had been the case during the third era.

In contrast to the residual losses, the costs of ultimate enforcement of informal rules can be considered as not being significantly affected by the switch from the third to the fourth era: As ultimate enforcement of informal rules is herein assumed to be realized by the members themselves in a spontaneous, distributed manner, it does not raise considerable costs for the centralized instance. There is, in this respect, no difference between the third and the fourth era.[26]

All in all, an "organization" with ever-changing membership statuses, with strongly blurred organizational boundaries and with cooperation having to be realized among locally dispersed "members" does not represent a *"close-knit group"* in the sense of Ellickson (1991, pp. 177f) anymore, and consequently, the approach of decentralized, spontaneous enforcement will be less effective within those "organizational" settings that characterize the fourth era.[27] Informal rules are less likely to develop at all, enforced less frequently and will lead to higher residual losses having to be accepted than it had been the case for the third era. We can thus assume increasing fixed and marginal costs, constant costs of ultimate enforcement and rising residual losses for the motivational meta-measure of informal rules to come along with the switch from the third to the fourth era.

Table 9.4: Motivation costs of formal and informal rules – Changes arising from the switch from era 3 to era 4.

Meta-Measure	Fixed Costs	Marginal Costs	Enforcement Costs (single case)	Residual Loss
Formal Rules	→	→	↗	↗
Informal Rules	↗	↗	→	↗

As a final aspect, we have to consider the possibilities of optimization. *Technical means* of optimization like those already mentioned in sections 7.2.2.1 and 7.2.3.1 –

[26] But see again note 24 on page 219. Of course, there is the possibility of interpreting member activities of spontaneous enforcement as hierarchical costs because of productivity losses. In this case, however, the respective losses would presumably also increase during the fourth era because of social enforcement requiring higher efforts. Alternatively, one could assume spontaneous enforcement to be less present, leading to higher residual losses – the effect for overall motivation costs of informal rules would basically be the same: They would increase together with the switch toward the fourth era.

[27] See also Posner (1997, p. 366), observing that norms are basically more effective within small groups, when the individual costs from being ostracized are high, and, in particular, when the respective group is static. Larger groups with constantly changing members are, in turn, *less* likely to develop and enforce effective social norms.

logging or notification mechanisms, for example – could unquestionably be used during the fourth era to lower motivation costs of formal and informal rules because of heightening the probability of nonconforming behavior to be recognized. But on the other hand, the strong asymmetries between the centralized instance and the individual member or those among the different members, respectively, together with the strong situation-dependence and the high level of uncertainty gives rise to a problem that does, in a way, resemble the concept of moral hazard already outlined in section 3.2.2.2: Even if the principal has substantial knowledge about the agent's actual behavior, the latter can ascribe this to the situational conditions necessitating or justifying the respective behavior and the principal, not being aware of these conditions, would in this case be unable to verify the adequateness of the agent's doings. For both, formal as well as informal rules, this results in lower possibilities for optimization through technical means and again, motivation costs will increase compared to the third era where situation-dependence was of significantly less relevance than it will presumably be in the fourth era. Nonetheless, technical means might very well still prove as a valuable mechanism for reducing information asymmetries regarding actual member behavior and thus for reducing motivation costs of formal and informal rules.

As a second possible means of optimization, *internalization* was mentioned for formal rules in section 7.2.2.1 as well as for informal rules in section 7.2.3.1. Basically, this mechanism rests upon individual feelings of guilt or pride, motivating members to enforce a certain formal or informal rule against themselves and, for the case of informal rules, to enforce an existing rule against other members, too. As, however, internalization is usually considered as being a long-term process[28], its effectiveness can also be assumed to decrease together with the transition toward the fourth era. The ever-changing social settings and the increasing dicontinuity of "membership" do not only imply a lower probability of informal rules to establish within an "organization" (see above) but also makes internalization of given formal *and* informal rules less likely to happen. "External" consultants or temporary workers, for instance, will usually not be "members" long enough to internalize a specific (formal or informal) rule regarding the use of external storage media or the access of public hotspots at all. While internalization of formal and informal rules could be used during the third era for reducing the necessary level of observation and the residual losses being present, its effectiveness will thus at least be limited by the specifics of the fourth era, too.

As we can see, the current and prospective developments toward a fourth era of interwoven computing heighten the hierarchical motivation costs of formal and informal rules in a multitude of ways. For both cases, the expected residual losses increase significantly because of the increased level of information asymmetries and uncertainty. The enforcement costs of formal rules will also rise and the development and change of informal rules will – if at all – only be possible at much higher costs than it has been the case during the third era. This, then, raises *general* doubts in the applicability of formal and informal rules as an effective and efficient instrument for enforcing a specific state of security-related member behavior in the fourth era. In the following section, these findings shall thus be fit into the summarizing representation of coordination

[28] See section 7.2.2.1.

and motivation costs that was developed at the beginning of this chapter.

Conclusion:
Besides the prognosis of a significant increase of coordination costs for the fourth era, our economically inspired model for cooperation with regard to security-related behavior also allows us to make well-founded statements about expectable developments of *motivation* costs – in particular with regard to those costs arising from formal and informal rules being used for motivation purposes. In the fourth era, these costs are likely to *increase* in a multitude of ways as a result of the local dispersion of "members", the increasingly temporary nature of "membership" and the continuous blurring of organizational boundaries. This raises serious doubts in the effectiveness and efficiency of the established motivational instruments of formal and informal rules within the context interwoven computing.

9.3 Implications

On the first pages of this chapter, we developed a highly abstract and generalized representation of the progress of hierarchical coordination and motivation costs over the different computing eras (see pages 203 to 206). According to this representation, cooperation with regard to security-related member behavior could be realized at low coordination as well as motivation costs during the first era of isolated computing, actually leaving most parts of the value realizable through computer usage to the organization. The introduction of shared mainframe systems and thus the switch toward the second era necessitated more extensive coordination efforts because of a multitude of factors and led to substantial coordination costs having to be borne. For the organization, this resulted in larger parts of the original value of shared mainframe systems being nullified by coordination costs while the amount of motivation costs did – due to the architectural means that were still used for motivation purposes – increase less significantly. The remaining overall value being derivable from the use of mainframe systems, however, must still have been larger than the respective remaining value of isolated systems as it would otherwise have been economically unreasonable for organizations to actually realize the shift from the first to the second era.

With the further shift from the second to the third era, however, coordination costs would have increased substantially for the hypothetical case of an organization trying to identify an aspired state of member behavior in a centralized, hierarchical manner in order to enforce this state by architectural means alone. Within our interpretation, these strongly increased coordination costs serve as explanation for the shift of prevailing information security practices that happened together with the transition from the second to the third era: Would organizations still have pursued the established approach of architecture-based motivation of members to behave in conformance to a predetermined coordination outcome, this would – compared to the second era – presumably have led to a *smaller* remaining value for third-era practices. Ultimately, this would have drawn the era-switch unreasonable and organizations should have kept

pursuing second-era practices instead. The use of formal and informal rules, however, allowed organizations to delegate ultimate decisions to the individual members and to cut hierarchical coordination costs substantially. Even if this led to higher motivation costs because of the specific characteristics of formal and informal rules, this must at least have led to a higher overall amount of remaining value than during the second era to make the switch economically reasonable for an organization (see figure 9.6).

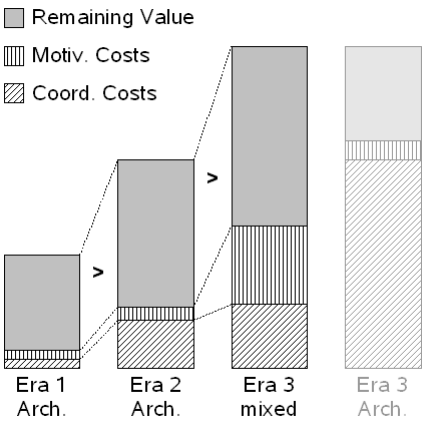


Figure 9.6: Change of cost-structures over different eras.

As already mentioned above, this interpretation is anything but proven from an epistemological point of view. Instead, it is nothing more – but also nothing less – than a plausible, reasonable and well-founded explanation for the changes of information security practices that actually happened in the past. If, however, we accept the substantial increase of *coordination* costs as plausible and reasonable explanation for the strongly increased relevance of formal and informal rules during the third era, and if we agree with the above remarks on the expectable development of motivation costs arising from formal and informal rules for the fourth era, then we might also come to the conclusion that the increase of these motivation costs could again render it unreasonable for an organization to actually switch from the third to the fourth era at all. Like it had been the case with coordination costs for the transition from the second to the third era, there is at least the possibility that the strongly increased *motivation* costs outreach the originally possible gains from shifting toward the fourth era of interwoven computing, leading to an overall *decrease* of remaining value and thus to the new paradigm not being adopted. It can thus be expected that at least in some cases, the heightened motivation costs will circumvent the profitable adoption of those principles and paradigms characterizing the fourth era and will thus lead to organizations adhering to the paradigms of the third era instead. Figure 9.7 illustrates this case of an overall decrease of remaining value resulting from strongly increased motivation costs that would restrain an organization from switching over to the fourth era – even if realizing this switch would originally allow for higher overall values to be generated.

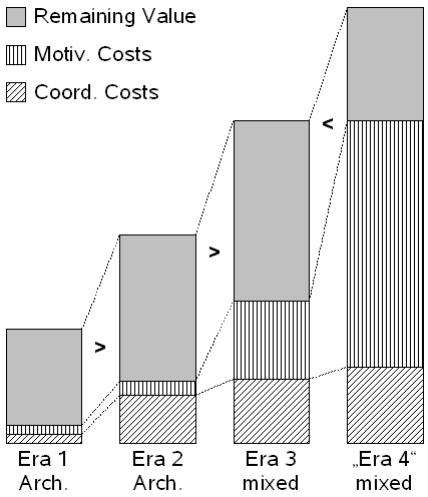


Figure 9.7: Change of cost-structure in "era 4".

What, then, does this imply for the future of information security practices?

First of all, security-related cooperation in the fourth era can unquestionably not be based on the "classical" principle of enforcing a well-defined state of aspired behavior through established architectural means alone. As the (hypothetical) hierarchical coordination costs having to be borne for such an approach would, starting from those of the third era, increase even further (see section 9.2.1 above), organizations would – on a generalized basis, of course – face the same drawbacks that already led to the change of prevailing information security practices together with the switch from the second to the third era. The main difference is that these losses are about to apply in an even aggravated form. Like it was the case for the switch from the second to the third era, this does not imply that explicit, detailed coordination and architecture-based motivation will not play *any* role at all during the fourth era, but as sole or primary approach for the fourth era, they are certainly unsuitable.

Formal and informal rules, in turn, were shown to presumably result in prohibitively high cooperation costs under the specific conditions of the fourth era, too. Like the substantial coordination costs of the "classical" approach based on architectural means would have drawn the switch from the second to the third era inefficient for a multitude of organizations, the motivation costs arising from formal and informal rules would – again, on a generalized basis – presumably have the same effect for the switch from the third to the fourth era. With the established approaches of information security alone, realizing the transition would in all likelihood lead to lowered overall benefits for organizations compared to the third era and consequently, organizations could be expected *not* to adopt the practices of the fourth era at all.

There are thus strong reasons to believe that the established principles and practices of organization-internal information security will prove inappropriate for future conditions and that we will – at least *also* – need new ones to properly address the specific

challenges arising in the fourth era of interwoven computing. But how shall these "new approaches" look like? What abstract principles should they be based on? And how could these principles be applied to information security "inside" the "organizations" of the fourth era? To answer these and many further questions, we shall in a first step identify some general requirements that have to be fulfilled by any "new approach" that should represent a viable alternative for realizing security-related cooperation in the fourth era. Subsequently, we will then discuss some possible future approaches on the basis of our economically inspired, abstract model.

Conclusion:
The significant increase of motivation costs having to be expected for the meta-measures of formal and informal rules give strong reasons to believe that these practices will not be viable anymore under the specific givens of the upcoming fourth era of interwoven computing. If security-related cooperation should be realized on the basis of existing and well-known practices alone, this would presumably deter a multitude of organizations from entering the fourth era at all.

Chapter 10

Future Directions, Alternative Approaches and the Regulatory Framework

Chapter 10

Future Directions, Alternative Approaches and the Regulatory Framework

This is not a dispute about whether planning is to be done or not. It is a dispute as to whether planning is to be done centrally, by one authority for the whole economic system, or is to be divided among many individuals.

– Friedrich A. von Hayek

After having identified a fourth era of "interwoven" computing in chapter 9 and after having discussed its implications on coordination and motivation costs, this chapter is devoted to the identification and discussion of possible future approaches for realizing security-related cooperation under the specific conditions of this new era. Altogether, we identified in this respect:

- An increased complexity of the whole cooperation process because of ever-changing, highly sophisticated and context-specific givens becoming relevant and because of ever-changing "membership" statuses.
- An increasingly high level of information asymmetries (between members and any centralized instance as well as among the individual members) regarding the specific context a certain behavior takes place in.
- An increasingly high level of information asymmetries (again, between members and any centralized instance as well as among the individual members) regarding the actual behavior of individuals.

Generally speaking, any novel approach to realize security-related cooperation has to take these changes into account. Furthermore, any approach still has to solve the tasks of coordination and motivation as outlined in chapter 3. Regarding the task of coordination, the specific givens of the fourth era imply that any viable future approach must provide a way to include extensive situational, context-specific information in the

coordination process and has to solve the problems of non-static "membership" statuses and high levels of complexity that were outlined in sections 9.1 and 9.2. Furthermore, this has to be realized without raising the substantial losses caused by established practices of hierarchical coordination. Motivation, in turn, will also have to be realized in a way that pays tribute to the specific givens of the fourth era. In particular, motivational instruments will have to address the same strongly increased asymmetries already influencing the coordination process which would otherwise presumably lead to prohibitively high motivation costs – at least for those means of motivation that feature a high level of violability (like the established formal and informal rules, for instance). Alternatively, one could also think of novel approaches to motivation that use instruments of low violability to make the respective information asymmetries less relevant in this respect.

There is a multitude of starting points for developing such novel approaches to security-related cooperation. As outlined in chapter 3, however, economic cooperation among different individuals can basically be realized through hierarchical and non-hierarchical approaches. Both of these modes can prove appropriate under specific conditions, but as demonstrated in section 3.3, a hybrid mode of cooperation, mixing centralized, hierarchical and decentralized, non-hierarchical approaches together, does in most cases represent the most efficient approach. Being aware of this advantage of hybrid modes, we will nonetheless maintain the distinction between hierarchical and non-hierarchical cooperation for demonstrative purposes and distinguish possible future approaches to information security on the basis of this dichotomy. However, none of the approaches discussed below will be of strictly hierarchical or strictly non-hierarchical nature. The dichotomy shall thus not be understood in a binary but in an "analogue" manner, distinguishing *rather* hierarchical from *rather* non-hierarchical approaches.

In a first step, we will thus outline some possible approaches to coordination *and* motivation that follow the established, primarily hierarchical practice in section 10.1. Subsequently, we will discuss some initial ideas for the future development of less (or even non-) hierarchical mechanisms for realizing cooperative security-related behavior in section 10.2. As we will see, both paradigms open up a wide variety of possibilities for originating new approaches and for refining existing ones – both on the basis of deliberate economic considerations.

As all these considerations are basically of a predictive nature, a certain vagueness will be inevitable throughout the following sections. Nonetheless, the discussion of some possible approaches for the depicted fourth-era challenges will at least allow us to lift the fog a bit and, not to forget, to structure our thinking about future strategies to information security in general. Let us, therefore, begin with those possible approaches that are rooted in the prevailing and well-known paradigm of hierarchical cooperation.

10.1 Some Hierarchical Approaches

From an abstract perspective, the rather hierarchical approaches that will be discussed in the following do all assume the existence of some kind of centralized instance(s)

being responsible for realizing security-related cooperation – including the subtasks of coordination and motivation – for the entire "organization". They are mostly based on established principles and practices but do expand those in one way or another to address the above-mentioned challenges. For each approach, we start with a general description of the approach itself and of the underlying abstract deliberations. We then discuss the expectable effects for coordination and motivation costs and close with some remarks on the implications of the respective approach.

10.1.1 Introducing Context to Traditional Architectural Means

The first hierarchical approach that shall be discussed here is the one that is most of all rooted in existing, "traditional" practices for realizing security-related cooperation. It primarily addresses the problem of increased motivation costs that will presumably arise from formal and informal rules in the fourth era and that will in all likelihood lead to substantial inefficiencies.

Due to the limited sustainability of formal and informal rules, it would possibly be advantageous to move backwards to second-era approaches based on architectural means of motivation to a certain extent. These raise significantly lower enforcement costs and do not – at least for the task of *motivation* – feature the serious effects arising from increased information asymmetries. Using architectural means would, however, require to limit *coordination* costs to a level that leads, together with their decreased motivation costs, to remaining derivable values that exceed those of the third era. Again, it would otherwise not be reasonable for an organization to actually carry out the transition from the third to the fourth era. How, then, could the specifics of the fourth era be introduced into a cooperation process that increasingly rests upon architecture-based motivation?

Regarding the employment of context-related aspects, Zhang and Parashar (2003), for instance, propose a system for *dynamic* role-based access control that prototypically introduces the concept of context to traditional mechanisms of access control. Basically, the proposed system does automatically change the effective role of a certain member whenever this member performs a context change. The system thereby limits or broadens the capabilities of the respective member depending on the context he or she is currently situated in.[1] Even if being originally aimed at environments of grid computing, such a mechanism would exactly allow what was abstractly delineated above: to introduce context to an architecture-based enforcement of cooperative behavior.

Comparable systems have been proposed in a multitude of variations. Al-Muhtadi, Ranganathan, Campbell, and Mickunas (2003), for instance, introduce a system for the field of ubiquitous computing that allows to formulate highly sophisticated, context-dependent access rules.[2] Ardagna, Cremonini, Damiani, di Vimercati, and Samarati

[1] It does, in this respect, make no significant difference whether the effective role of a certain member is changed or whether the capabilities attached to an unchanged role vary depending on the context. Note, however, that the concept of roles does, in turn, already represent a generalization itself (see section 6.2.1).

[2] The authors themselves use the term of *"[c]ontext-aware security policies"* (Al-Muhtadi et al. 2003,

(2006) introduce different aspects of location to established mechanisms of access control, Chandran and Joshi (2005) use location and time to augment established models and Covington, Long, Srinivasan, Dev, Ahamad, and Abowd already provided an exhaustive formal model for context-aware access control mechanisms in 2001.

Many further examples could be given, but all the mentioned approaches have, from the abstract perspective pursued herein, in common that they concentrate on the *motivational* aspects of cooperation. We will therefore discuss the cost implications of introducing context to architectural means "from the back to the front", starting with *motivation* costs.

10.1.1.1 Motivation

As mentioned above, using architectural means for motivation purposes could overcome the problem of substantial motivation costs arising from formal and informal rules in the fourth era. Instead of either having to accept these significant motivation costs or abstaining from fourth-era practices at all, an organization could use, for instance, a context-enabled architectural access control mechanism for motivating a certain individual to behave in a specific manner whenever residing within a specific context. This would make any additional observation efforts obsolete and eliminate the risk of residual losses arising from increasing amounts of nonconforming member behavior. An issue like the processing of a highly sensitive internal document outside of the organization's building could, for example, be controlled by means of a location-aware access control mechanism. Due to the high strictness and low violability of architectural means, no significant losses arising from rule-breaking behavior had to be expected and presumably, the overall motivation costs would be reduced as compared to those arising from formal and informal rules in the fourth era.

Furthermore, the use of context-enabled architectural means could also overcome the problem of possibly ambiguous scopes of formal and informal rules arising from ever-changing membership statuses. As architectural means are self-enforcing, any problems arising from individuals not being aware of a certain rule or being in doubt what formal and informal rules to follow in a given situation could be overcome and ex-post punishments with regard to "externals" would not have to be considered anymore.

On the other hand, not all specific characteristics of the fourth era can easily be addressed by context-sensitive architectural means. Take, for instance, the fact of increasingly relevant and ever-changing "external" relations within an organizational setting that strongly rests upon service-oriented architectures (SOA) and software-as-a-service (SaaS) and that makes extensive use of outsourcing formerly internal processes to external contractors: Using architectural means of motivation under such conditions would require these means to be applied to any single technical entity (temporarily) belonging to the scope of the "organization". Any single external contractor as well as any externally sourced service had to be equipped with the respective architectural means. The intuitively felt skepticism about the realizability of such a scheme is also backed by abstract considerations regarding the cost structures of architectural

p. 496), which shall not be confused with the understanding of a security policy pursued herein.

means: Even if architectural means raise low costs of ultimate enforcement and low residual losses, they are on the other hand subject to high initial "fixed" costs.[3] These setup costs, in turn, had to be shouldered initially for any external instance that should be subject to the respective motivational means, even if the affiliation to the organization is of temporary, short-term nature alone – like it could be the case for an external service that is only used for a single project, for example. The significant economies of scale provided by architectural means which make them highly efficient within rather static settings can thus not be benefited from with ever-changing "external" relations being increasingly relevant for the whole process of cooperation.

Overall, context-enabled architectural means could provide significantly lower *motivation* costs than the highly violable formal and informal rules under some of the the specific givens of the fourth era. Especially with regard to the problems of short-term memberships and of ever-changing context and strong information asymmetries that arises from vast member mobility, the introduction of context to architecture-based motivational means could thus serve as economically viable alternative. Other fourth-era phenomena like the increased relevance of temporary external relations, however, will presumably not be adequately addressable by mechanisms like those discussed here. In the latter case, the high initial setup costs of architectural means will presumably render their application inefficient.

Besides motivation, however, any mechanism for realizing cooperative behavior also has to solve the task of *coordination*. We shall thus discuss this aspect in brief, too.

10.1.1.2 Coordination

Generally speaking, any mechanism for realizing motivation on the basis of architectural means in the described manner would require some state of *context-specific* aspired member behavior to be formally specified at least in some detail in a preceding coordination process. This would in turn raise the same problem of prohibitively high coordination costs that was already discussed for the switch from the second to the third era. The availability of context-dependent architectural means alone would still be worthless without the existence of a context-specific state of aspired member behavior that should be codified into these means and be enforced through them.

Unquestionably, any attempt to reach such a coordination outcome based on extensive consideration of contextual aspects would lead to immense coordination losses. All those factors already discussed in section 9.2.1 would apply in this respect, too: the increased need for repeated reconsideration (or the need to consider any relation with regard to a multitude of possible contexts, respectively), the higher complexity of any single decision, the high level of uncertainty being faced by any centralized coordination instance and the impossibility of anticipating every contextual aspect that could possibly become relevant in the future. These factors must simply lead to any highly detailed and formalized coordination outcome entailing significant coordination costs.

[3] These costs arise from the need for setting up the technical solution, for integrating it into the existing environment, or for eliminating incompatibilities, for example. For further details, see section 7.2.1.

In contrast, any approach that does *not* explicitly take context-specific aspects into account can be understood as assuming one and the same context for all relations being considered. For instance, the access to and revision of a specific, highly sensitive, organization-internal document by a specific member would always be evaluated in the same manner, always leading to the same aspired behavior ("should" or "should not" access and revise) as outcome of the coordination process, regardless of whether the member currently resides in one of the organization's internal office rooms, in a private hotel room, in the lobby of a hotel or even on a train. Abstaining from these separate considerations for different contexts would thus in all likelihood result in significant costs of maladaption.

Using the terms introduced in section 6.2.1, we can interpret this non-consideration of context-related aspects as just another case of *over-generalization*. At least implicitly, all possible contexts would be lumped together and treated in a uniform, singular manner. The respective costs of over-generalization, in turn, cannot be overcome completely because of any attempt to consider context in detail would rise other significant coordination costs, too. But as also delineated in section 6.2.1, an *intermediate* level of generalization does usually provide more advantageous cost structures than the two extreme cases of "no generalization" being present and "complete generalization".[4] Realizing such an intermediate level of generalization for context-related aspects thus promises to reduce overall coordination costs as compared to practices that do not differentiate over different contexts at all and those that do so extensively. An abstract graphical representation is given in figure 10.1.

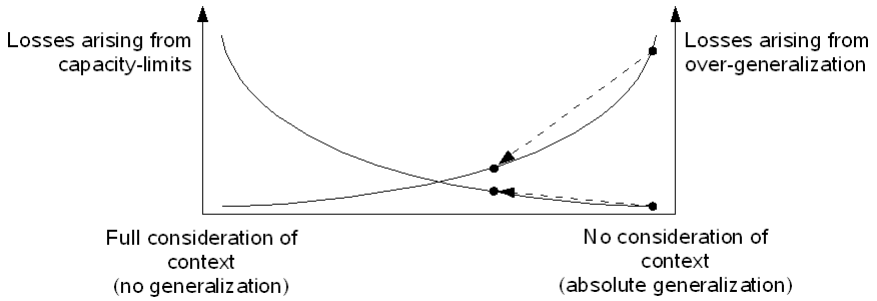


Figure 10.1: Cost trade-off between extensive consideration and non-consideration (over-generalization) of context.

Using the above-mentioned mechanisms for introducing context at an intermediate level of generalization could thus at least reduce the coordination costs that would otherwise have to be borne by an organization within a paradigm of architecture-based enforcement of cooperative member behavior. Besides generalization, other means of optimization like the integrated solutions for expanding the effective capacity limits that were already introduced in section 6.3.1 could of course also be applied. By all those approaches, a centralized coordination instance could minimize overall

[4] See, in particular, the graphical representation given in figure 6.3 on page 125.

coordination costs and come to more or less efficient coordination outcomes that could then be enforced through the above-mentioned, "context-aware" architectural means. The resulting overall coordination costs, however, will still be higher than those of the second era and, of course, higher than the coordination costs from the third era, which were significantly reduced through extensive delegation of ultimate decisions to individual members.

All this notwithstanding, there still seems to be no viable approach for addressing the *really* profound changes characterizing the fourth era within a hierarchical coordination process that should produce architecture-enforceable outcomes. Compared to the implications arising for the coordination process from different organizations being complexly interwoven with each other and from externally sourced services being used to realize "internal" functions in constantly changing combinations, the mere introduction of member mobility is relatively simple. A hierarchical, centralized coordination process that should evaluate a complex network consisting of a multitude of relevant entities (technical ones as well as individuals) – which, in turn, belong to a multitude of different organizations – had to solve a completely different class of complexity, non-predictability and need for repeated reconsideration. The multitude of ever-changing instances being interconnected with each other in an ever-changing manner will presumably result in a formerly unknown complexity of the coordination task.[5]

Furthermore, a profound evaluation of external services and of other organizations is impossible without knowing the inner workings and structure of these services or organizations. This knowledge can, however, also not be assumed as given to a centralized coordination instance. Realizing centralized coordination with formalizable outcomes under such conditions is not impossible, of course, but we can expect substantial losses to arise from such a process. Even an intermediate level of generalization would – due to the much more complex interrelations having to be regarded – presumably result in an extremely maladapted coordination outcome.

Like it was already the case for the task of motivation, the introduction of context to architectural means might thus allow to realize cooperation with a certain efficiency for the case of (part-time) members of a well-specified organization acting within locally dispersed and constantly changing environments. This is, however, not the case for further and more radical changes characterizing the fourth era like, in particular, the increasing use of external services or the transition toward strongly interwoven "organizations". Realizing explicit and centralized coordination would for these cases presumably still result in enormous losses.[6]

[5] For a less abstract perception of the same aspect, see also Denning and Hayes-Roth (2006), noting that *"centralized decision making does not work in large federated networks"*.

[6] Just for the sake of completeness, it is interesting to note that none of the models for introducing context to traditional architectural means outlined above does explicitly address the *coordination* problem of how the proposed, highly complex access rules are to be developed at all.

10.1.1.3 Implications

At least for some aspects of security-related cooperation among the "members" of a fourth-era "organization", the overall cooperation costs of the architecture-based approach discussed here could be lower than they would presumably be with formal and informal rules being employed. Especially with regard to member mobility across varying but more or less generalizable contexts, coordination could, with an intermediate level of generalization, be realized with a certain efficiency. Instead of having – compared to the third era – to delegate even more ultimate decisions to individual members by means of formal and informal rules, this would allow to decide over collectively aspired member behavior *ex-ante* for more cases.[7] This heightened amount of relations being ultimately decided over ex-ante would, in turn, more often permit to motivate members to behave in conformance with these ex-ante decisions through architectural means than it would be the case without introducing context to the centralized coordination process.

Regarding the task of motivation, context-enabled architectural means will for the same cases – due to the significantly increased motivation costs of formal and informal rules – presumably get back the relative cost advantages they already held during the second era. Even if the overall costs of architecture-based cooperation will still increase as compared to the overall costs of the mixed-approach from the third era, this increase will in all likelihood still be smaller than it would be with established practices being used without further changes.

Especially with regard to member mobility across varying but more or less generalizable contexts, the remaining overall value for the organization could possibly even exceed that of the third era, making the adoption of the respective practices economically reasonable for an organization. Compared to the third era, it would then be advantageous to apply the "traditional", architecture-based approach to *more* cases again – provided that the respective solutions actually allow to realize an intermediate level of context-dependent coordination and motivation. The resulting cost structure is illustrated in figure 10.2. At least to a certain extent, mechanisms based on context-enabled architectural means might thus represent a less – but still highly – inefficient alternative to formal and informal rules for realizing security-related cooperation under the conditions of the fourth era.

The more profound changes defining the fourth era, however, will presumably not be adequately addressable by context-enabled architectural means. As outlined above, aspects like the strong interwovenness of "organizations", the extensive use of "external" services and "external" contracting could not be properly addressed – neither with regard to coordination, nor for the task of motivation. For these aspects of the fourth era, the approach of context-enabled architectural means thus does *not* provide a viable alternative for realizing security-related cooperation as it is strongly unlikely that their use would allow to realize remaining overall values that supersede those of the third one. Consequently, it would, the existence of context-enabled architectural means notwithstanding, still be more valuable for an organization to persist in third-

[7] Nonetheless, the overall coordination costs of doing so will in all likelihood supersede those of the second and the third era.

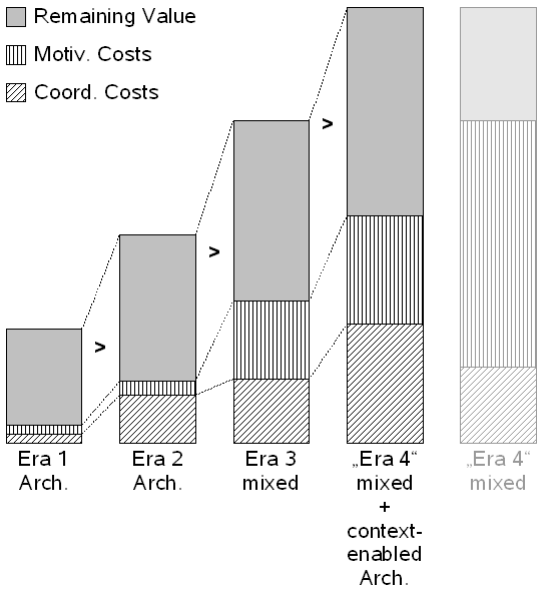


Figure 10.2: Possible cost advantage of context-enabled architectural means over established approach.

era practices than to carry out the transition to the fourth era.

The latter-mentioned developments are, however, unquestionably more crucial for the recognition of a new "era" than the mere introduction of member mobility which could also be seen as a first phenomenon of "post-third-era" developments gaining momentum at all. Rather than actually being an answer to the real fourth-era requirements, the introduction of context to architectural means of motivation and to the corresponding coordination activities thus represents a refinement of the mixed, third-era approach. It allows a higher amount of relations to be decided ex-ante without raising extraordinary costs of maladaption. Especially for cases involving high asymmetries with regard to actual member behavior – like the "post-third-era" developments toward higher member mobility within an otherwise mainly unchanged setting – significant motivation costs could be eliminated without raising too high coordination costs. The whole gamut of the fourth era, however, does not seem to be solvable through the introduction of context-enabled architectural means.

Conclusion:
The introduction of context to architectural means of motivation might allow to overcome the problem of increasing motivation costs arising from formal and informal rules for scenarios involving member mobility within otherwise mainly unchanged organizational settings. The really profound changes constituting a new era, however, can presumably not be addressed properly by such mechanisms. In particular, the extensive coordination costs would in all likelihood still prevent an organization from actually realizing the transition from the third to the fourth era.

10.1.2 Expanding Logging Mechanisms to Contextual Aspects

As mentioned above, the overall cooperation costs of the fourth era would increase significantly within the established "mixed" approach from the third era including strong delegation because of strongly heightened motivation costs. Instead of – as suggested in the previous section – abstaining from the use of formal and informal rules at all in order to solve this problem, we could also search for ways to use these formal and informal rules in a more efficient way. Such an approach shall be discussed in the following and again, economic theory shows us the way toward this approach.

The fact of motivation being costly to realize as well as an inherent need for delegating a certain amount of decision-making to individual subordinates are, for instance, also the basic assumptions of the whole field of agency theory. As already outlined in section 3.2.2, a principal-agent relation is, besides the general need for a certain amount of delegation, mainly characterized by two further properties: by strong information asymmetries existing between the principal and the agent and by diverging interests of those two players. Such principal-agent relations thus show strong similarities to those situations arising for the case of established third-era approaches of delegation being maintained during the fourth era.

On the other hand, agency theory knows various possibilities for counteracting the resulting drawbacks to a certain extent. In particular, agency theory subdivides information asymmetries relevant for ex-post enforcement[8] into those regarding *hidden information* and those of *hidden action*, which both increase the risk of opportunistic behavior taking place.[9] In the latter case of hidden action, the principal is unaware of the agent's actual behavior. The agent, in turn, is very well aware of this fact and

[8] The complementary *ex-ante* information asymmetries regard the general quality of the agent. They are relevant *before* an agent is employed by the principal and thus before delegation takes place at all. This might be of particular importance for the case of outsourcing where information asymmetries are nowadays already counteracted to a certain extent through mechanisms of due diligence, for example. Agency theory, however, also knows other mechanisms for overcoming these *ex-ante* asymmetries. In particular, the concept of *"meaningful signals"* as developed by Spence (1973) could serve as rich source for further discussions in this field – regarding possible outsourcing providers *as well as* individual would-be "members". These *ex-ante* asymmetries and the possible mechanisms for reducing them will be of certain relevance for the rather non-hierarchical approach that will be discussed in section 10.2.2 below. They shall, however, not be discussed in more detail here.

[9] See, for instance, Furubotn and Richter (2005, p. 200).

therefore has an interest in behaving opportunistically – that is, against the principal's interest – as he does not have to fear any adverse consequences to arise from such unobserved opportunism. It is this problem of hidden action that all the logging mechanisms already mentioned above[10] try to to solve to a certain extent. If any action of the agent – connecting to a hotspot, for instance – is logged, ex-post asymmetries regarding actual behavior can be significantly lowered and the agent does at least have a lower interest in behaving opportunistically because of his formerly "hidden action" becoming "visible" to the principal.

This would, however, still leave the problem of moral hazard unanswered. This refers to the fact of the principal not being able to verify the *adequateness* of the (logged and thereby revealed) actual agent behavior because of hidden information regarding the context the behavior had taken place in. Due to this hidden information and due to the agent's awareness of this information asymmetry, the agent has the possibility to ascribe bad outcomes to adverse conditions and good outcomes to adequate behavior. The principal, in turn, has no possibility to assess the substance of such arguments. Another possible approach for lowering motivation costs under the conditions of the fourth era would thus be to lower these information asymmetries, too.

If, for example, logging could be expanded to a multitude of further aspects and, in particular, to those also informing the principal about the specific context a certain behavior had taken place in, this would counteract the problem of moral hazard. The principal would then better be able to estimate the adequateness of the agent's behavior ex-post and the agent, being aware of this fact, had fewer possibilities of (untruly) ascribing bad outcomes to adverse circumstances and not to inadequate behavior of his own. One possible answer that could be derived from agency theory would thus be to counteract the problem of moral hazard by extensive logging of individual behavior which also includes as many contextual information as possible. In particular, this refers to those contextual information that allow the principal to better answer the question for the adequateness of a given member behavior: location, external connections being made, further activities in the same time frame, etc. would be possible candidates.[11]

How, then, would such mechanisms presumably affect the costs of coordination and motivation with regard to security-related behavior?

10.1.2.1 Coordination

Generally speaking, the coordination costs would in this approach be unaffected as compared to the hypothetical case of third-era practices being maintained without further change. As already outlined in section 9.2.1, the characteristics of the fourth era with the even increased information asymmetries give good reasons to keep on delegating a multitude of ultimate decisions to the individual members. Different

[10] See sections 7.2.2.1, 7.2.3.1 and 9.2.2.

[11] This does, of course, raise the issue of such practices possibly not being allowed for legal reasons or because of other constraints that could be assigned to the "regulatory framework" (see section 2.2.4). Let us nonetheless ignore such – unquestionably warrantable – objections for the moment. We will return to this aspect later in section 10.1.2.3 and, more extensively, in section 10.3.

from the first approach discussed in section 10.1.1 above, this high level of delegation would be retained, leading to a similarly low level of *coordination* costs having to be borne by the organization.

10.1.2.2 Motivation

Regarding the costs of motivation, logging mechanisms that include contextual information would represent a possibility to limit the problem of moral hazard and thereby to reduce the costs arising as residual losses for those motivational mechanisms following the approach of ex-post enforcement. Most obvious is the possibility for applying extended, context-enabled logging to formal rules which are enforced through ex-post sanctions by a dedicated, centralized instance. If there is, for example, a formal rule saying that WLAN-hotspots should only be connected to in an insecure manner when doing so would unquestionably be in the organization's interest, contextual information about the websites or other services being accessed over such a connection would allow for better evaluations about whether this interest of the organization actually existed or not. The same would be possible for formal rules saying that sensitive, organization-internal documents should only be opened and edited when residing within specific (types of) locations.

Applying the same mechanism to the other ex-post method of informal rules would, in general, also be possible. But as informal rules are enforced among the individual members themselves in a decentralized manner, this would require the respective contextual information (and the information about the actual behavior, of course) to be made visible ex-post to all those other members that should eventually engage in the enforcement of informal rules.[12] Possible objections regarding the legitimacy of such mechanisms left aside and assumed that such mechanisms could be appropriately realized in the future, this would then allow individual members to better evaluate the appropriateness of their co-members' behavior based on that contextual information and would thereby limit the risk arising from moral hazard for the field of informal rules, too. In the end, this would also lower residual losses and thus also limit the motivation costs arising from informal rules under the conditions presumed for the fourth era.

Generally speaking, context-enabled logging mechanisms would, in a sense, re-establish the costs-structures of motivation known from the third era. At least some of the specific characteristics of the fourth era – in particular, the strong location-dependence and the resulting risk of moral hazard – could be counteracted by such mechanisms. The already established mechanisms of ultimate enforcement could still be applied to a certain extent and ultimately, overall motivation costs could be reduced as compared to the hypothetical fourth-era state outlined in section 9.2.2. In the end, ex-post motivation could possibly be realized at reasonable overall costs again and ultimately, the organization could possibly derive higher remaining overall values from fourth- than from third-era practices.

[12] Think, in this respect, of the consciously overdrawn idea of *"posting a list of all users who accessed pornography in the last week on the company notice-board"* mentioned by Povey (2000, p. 44), for example.

Other aspects of the fourth era, however, will in all likelihood only hardly be addressable by the mechanisms of extended logging described herein. For instance, the problem of ambiguous scopes of formal and informal rules would last unsolved. Even with extensive logging of contextual information, an external consultant might still be uncertain about the respective rules having to be obeyed in a given situation. Similarly, possible problems with regard to the exertion of ex-post punishments to "externals" or temporary "members" would also remain.

Furthermore, and like it was the case for the context-enabled architectural means discussed above, applying context-enabled logging to an increasing number of external relations with outsourcing partners or SOA-/SaaS-providers will presumably prove elusive. On the one hand, this would lead to the same problem of *any* such external party having to adopt a certain approach or solution that was already mentioned for context-enabled architectural means and on the other hand, such logging had to be realized in a trustworthy manner that cannot be forged by the party being monitored as the principal would otherwise not be able to rely on the logs.

And finally, extensive logging of actual behavior and of the respective context would lead to huge amounts of information that could hardly be overseen in its entirety. Even if the respective data were only examined *after* an adverse event having occurred, this would result in higher costs having to be borne for investigative activities and the act of estimating the adequacy of a certain behavior in a given context would become more sophisticated – and thus more costly – too. The motivation costs of this approach can therefore be expected to be higher than those of the third era.

Altogether, extensive logging of contextual aspects accompanying a certain member behavior would possibly represent a viable alternative to context-enabled architectural means for some aspects of the fourth era. Especially for the mere challenge of counteracting moral hazard under conditions of relatively static membership statuses and of the adequacy of member behavior being strongly context-dependent, it would still allow the established approach of delegating ultimate decisions to be retained instead of abandoning it in order to use architectural means of motivation. For other phenomena characterizing the fourth era, however, context-enabled logging will presumably prove inadequate.

10.1.2.3 Implications

From a strictly economic perspective, this leads us to the following structure of cooperation costs under conditions of extensive and context-enabled logging: Coordination costs would stay unchanged as compared to the default cost structure for the fourth era forecast in section 9 as the organization would still make strong use of delegating ultimate decisions to the individual members who are much more familiar with the specific givens of a certain situation. Motivation costs, however, would very well be affected by mechanisms of context-logging. As the respective enforcement instance of formal rules (and possibly of informal rules, too) would be able to make better evaluations of the appropriateness of a certain behavior, the risk arising from moral hazard could be lowered and we could expect a significant reduction of residual losses.

As outlined above, this reduction of motivation costs can, however, only be applied

to some of the specific characteristics of the fourth era. Aspects like the increased relevance of outsourcing or of externally obtained services will presumably be as hardly addressable as the problem of possibly ambiguous scopes regarding the applicability of formal and informal rules. Even more than the context-enabled architectural means from the previous section, the approach of context-logging would thus be limited to settings that do not differ much from those of the third era. Its application will presumably be most valuable for organizations with comparably static memberships and with these members acting within ever-changing contextual environments.

For these settings, however, extensive logging including contextual aspects could very well reduce information asymmetries between principal and agent not only with regard to hidden action but also to hidden information (see above) and could thereby diminish the losses that had otherwise to be expected because of the problem of moral-hazard. In the end and still from a strictly economic perspective, this would allow formal and informal rules to be applied at lower motivation costs and possibly enable the organization to realize higher overall remaining values than it was the case during the third era, thereby rendering at least some fourth-era practices economically reasonable. Figure 10.3 visualizes this case.

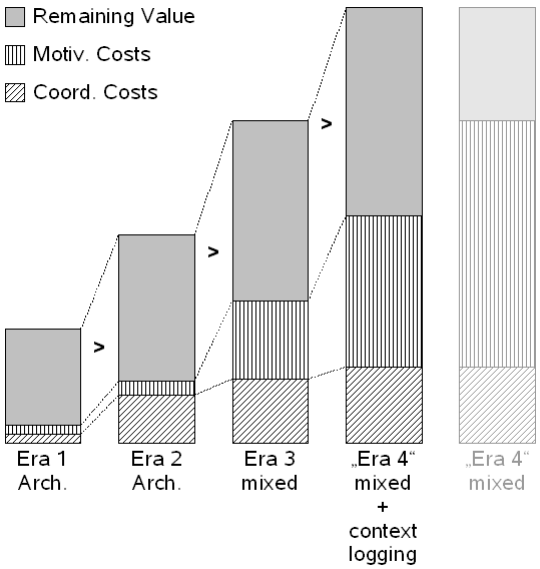


Figure 10.3: Possible cost advantage of context-enabled logging mechanisms over established approach.

Even if the approach of extended logging of user activities *and* of additional contextual information would thus allow to counteract the problem arising from increased information asymmetries together with ex-post motivation to a certain extent, the respective reduction of motivation costs would only be realizable at a high price: As outlined above, applying this mechanism to the enforcement of formal rules would not

only require to log nearly *any* user activity in order to be able to review the respective behavior later, but it would also necessitate extensive logging of a multitude of further information like the location the respective user sojourned at a given time, what files he accessed during the same time frame or what phone calls he made in the same evening. One could think of a multitude of further aspects also being relevant for the later evaluation of the situational adequateness of a certain behavior and in the end, we would result in a situation where the user is monitored more or less completely by the principal just to reach a lower level of motivation costs or, respectively, a higher level of information security. For the case of trying to extent the same approach to informal rules, this would require a multitude of comparable facts not only to be revealed to some dedicated, centralized instance but rather to be published among *all* members ex-post in order to allow for the usual acts of spontaneous enforcement to be actually realized.

Such practices would, however, violate the given regulatory framework in a multitude of cases. Consider, as a relatively obvious example, legal privacy regulations: Even if promising a certain business value for organizations, extensive logging of user activities and of contextual information would be prohibited by most privacy laws because of the personal rights of the members being seriously affected and because of these personal rights being valued higher than the obtainable efficiency gain under most legal regimes.[13] The approach of extensive logging of context for the purpose of later ex-post evaluations is thus a prime example for a practice that would presumably allow for a more efficient cooperation process inside organizations but that would at the same time go beyond the boundaries given by the regulatory framework and that could thus actually *not* be profited from. We will discuss some further aspects of the regulatory framework later in section 10.3 but for now, we will proceed with another rather hierarchical approach for realizing security-related cooperation during the fourth era.

Conclusion:
The expansion of logging mechanisms to contextual aspects would reduce information asymmetries not only regarding an individual member's actual behavior ("hidden action") but also those asymmetries concerning the specific information an agent was situated in ("hidden information"). It would thereby allow for better ex-post evaluations of the adequateness of an individual member's behavior and reduce the motivation costs arising from the problem of moral hazard. Even if thus representing a possible means for heightening the efficiency of cooperation for some specific aspects of the fourth era, the use of such mechanisms would nonetheless be prohibited by the regulatory framework in the majority of cases.

[13] It shall herein in no way be put into question whether these regulations are "right" or not, even if some law & economics scholars like Posner (1978) would probably value efficiency over the personal rights here, too. Such discussions, even if highly inspiring from an academic point of view, would go far beyond the actual subject of *this* work and shall thus not be pursued further herein.

10.1.3 Combining Violability and Architectural Means

The third rather hierarchical approach that shall be put up for discussion here does, like the other ones, primarily addresses the substantial motivation costs of formal and informal rules arising from the strongly increased information asymmetries of the fourth era. As illustrated in section 9.2.2, these asymmetries lead to considerable residual losses having to be expected because of the *high* violability of formal and informal rules and because of individual members facing a lower probability of rule-disobedience being detected. On the other hand, it is just the low strictness of formal and informal rules – as compared to traditional architectural means like access control lists – that qualified these motivational instruments for realizing the delegation of ultimate coordination decisions to the individual members who are much more familiar with the context-specific givens of a certain situation.

Instead of giving back this possibility of profiting from individual, situation-dependent knowledge by means of delegation (like it was suggested in section 10.1.1), one could also try to preserve this possibility while (differently from the approach discussed in section 10.1.2) at the same time still profiting from the specific characteristics of architectural means. This would, however, require to combine architectural means of motivation with the fundamental violability provided by formal and informal rules.

In fact, this is exactly what various existing approaches for changed access control paradigms are about. Povey (2000), for instance, suggests a model of *"optimistic security"* which follows the idea of technically enforced access rules to be consciously exceedable by the respective users. The model explicitly refers to *"situations where the ability to relax normal access rules can become critically important"* (p. 40) and is, generally speaking, suggested to be realized through the introduction of a third mode for defining access privileges. This third mode should, besides "granting" and "not granting" access to a certain resource, essentially prohibit access but allow the user to explicitly disobey this prohibition in a well-defined process. This well-defined process, in turn, is suggested to include a *temporary* logging mechanism, an obligatory audit trail and some kind of *"retrospective"* (or, within our terminology, ex-post) enforcement to be realized after the adequateness of user behavior has been evaluated by the administrator (or the centralized instance, respectively).[14] As suggested by Povey, such mechanisms should of course not be applied where the risk arising from rule exceedance is high as compared to the resulting benefits. But where the risk is comparably low and where the possibility to explicitly override access rules provides significant advantages over traditional methods, it could very well be applied to allow for better overall outcomes.

Even if originally aimed at various cases of emergency, this model could also be employed for other, less dramatic contexts. Rissanen, Firozabadi, and Sergot (2006), for example, also refer to the fact that *"unanticipated situations occur all the time"* (p. 312) and propose an access control mechanism that does, within certain boundaries, allow overriding. In particular, their model also rests upon *three* different kinds of privileges that can be given to a user with regard to a certain resource: *"permitted*

[14] Furthermore, Povey (2000) also suggests mechanisms for automatic rollbacks to previous system states. These shall not be discussed further herein.

access, denied access and access possible with override" (p. 315). Overriding within the third mode is realized only after notification of the user and after this user had explicitly approved his will for making use of his overriding capabilities. Like it is the case for Povey's model described above, any such case of a user having consciously and explicitly overridden an access rule is logged and reported to some instance for realizing ex-post evaluation of adequateness and for exerting punishments if necessary.

Due to these characteristics, overridable access controls might be abstractly understood as an alternative enforcement mechanism for the meta-measure of formal rules. Like these, overridable access mechanisms know behaviors that are always prohibited, other activities that are always allowed and, most importantly, a certain amount of cases where the admissibility *depends*, to some extent, on further aspects that have to be evaluated ex-post. On the other hand, overridable access controls need comparatively *explicit* rules to be effectively codified into and represented by them. Again, this suggests a certain relation to formal rules while informal rules are typically less suitable for being translated to the mechanisms discussed here.[15] Assume, therefore, that overridable access control mechanisms are used for enforcing "traditional" formal rules in an alternative manner.

This approach has recently been seized in various ways. Ferreira, Cruz-Correia, Antunes, Farinha, Oliveira-Palhares, Chadwick, and Costa-Pereira (2006), for example, actually applied such a model of overridable access control within a health-care context, Padayachee and Eloff (2007) suggest to enhance the paradigm by means of usage control, and Zhao and Johnson (2008) made a first attempt to analyze the approach within a strictly formal model of mathematical economics. At least at first sight, the concept of exceedable access control mechanisms seems like an interesting alternative for solving our fourth-era challenges and its implications for coordination and motivation costs shall thus be discussed in the following.

10.1.3.1 Coordination

Consistent with our above-stated comprehension as alternative enforcement mechanism for formal rules, overridable access control mechanisms unquestionably represent another way for delegating ultimate decisions to individual members. Even if users are in a first step hindered from accessing a certain resource in the intended manner, this is not done in an absolute way. Instead, the user is informed that the respective access is essentially prohibited but that this prohibition can be overridden, given that the user agrees to some specific conditions. The user himself can then decide whether he wants to make use of this possibility and, if he wants to do so, has to explicitly confirm this decision. After this acknowledgment, he is granted the intended kind of access to the specific resource.

Within our WLAN-example, this approach could, for instance, be realized by a technical solution that initially blocks any insecure network connection that could not be protected by a VPN from the outset. But instead of blocking access completely, the solution could raise a popup window informing the user about the initial restriction,

[15] One might, however very well speculate about possible approaches for doing so.

possibly showing some explanations for this blockade or even the formal rule being enforced by the restriction. The user would then be allowed to exceed the restriction by pressing a button labeled "access, anyhow", for instance. Furthermore, the solution could also require the user to give some justifications for overriding. Only after the user has explicitly declared his intention to consciously override the restriction, access to the unprotected network would then be granted while at the same time, extensive logging would be initiated for the purpose of later ex-post inspections. The ultimate decision about actually overriding a restriction, however, would rest with the individual user who is in the best position to decide whether doing so is reasonable in a certain situational context or not.

A certain amount of centralized coordination would nonetheless have to be present. In particular, the fact of *three* different modes being attributable to any kind of access to a certain resource requires these modes to be actually set and, in particular, to be coordinated in advance. There will always be some cases where overriding should not be possible at all and there will also be cases for which access is to be granted by default. On the other hand, the addition of a third mode also allows the coordination process to be more differentiated. In the end, this will presumably lead to coordination costs comparable to those of classical delegation realized by means of formal and informal rules.

These will, however, presumably be heightened to a certain extent because of the additional necessity for consciously ascribing one of the three modes to any kind of access with regard to any resource instead of – like it is assumed for established approaches of delegation – generally allowing the user to perform most kinds of access to most resources. Any such attribution, in turn, would presumably lead to a certain amount of misallocation and thus to some additional coordination costs.[16] This leads us, as compared to the forecast level from section 9.3, to slightly increased coordination costs within the approach of overridable architectural means.

10.1.3.2 Motivation

Motivation, in turn, is in this model realized on an ex-post basis. Like it is the case for motivation being realized through formal and informal rules, users basically have the possibility to disobey a certain rule within some given boundaries. Like the meta-measures of formal (and informal) rules, this mechanism thus features a high violability which is – and this is the crucial difference from established violable means – not an *inherent* property of the meta-measure but rather *artificially* built into it. As outlined above, this is done in order to being able to delegate a certain amount[17]

[16] Some kinds of access might, for example, "wrongly" be attributed as "always prohibited" or "always allowed". Reconsider, in this respect, the notes on optimization approaches like generalization from section 6.2.1.

[17] Note that "traditional" formal and informal rules are in most cases also used *together* with architectural means in the mixed approach pursued nowadays. Like in the approach discussed here, the violability of formal and informal rules does thus only cover a limited area while for some resources, a high violability is usually deemed inappropriate and architectural means are used in these cases. The violability of formal and informal rules is thus only granted within some given boundaries, too.

of ultimate decisions to the individual members who are much more familiar with the specific situational givens and can thus be assumed to be – at least for some cases – able to make better decisions than a centralized instance could in advance.

The fact of decisions being delegated to the individual member does, in turn, again necessitate mechanisms that motivate individual members to make decisions that actually represent the collective interest of the organization and that do not serve purely opportunistic goals. As mentioned above, explicitly overridable access restrictions can be seen as alternative mechanism for enforcing formal rules in this respect. For "traditional" enforcement methods of formal rules, however, the motivational strength (and thus, the amount of residual losses having to be expected) is determined by the likelihood of rule-breaking behavior being actually discovered by the enforcing instance and on the severity of possible sanctions. As outlined repeatedly in the preceding sections and chapters, it is primarily the decreasing probability of detection that, resulting from increased ex-post asymmetries, will in all likelihood decrease the efficiency of formal (and informal) rules during the fourth era.

This is different for the approach discussed here. As it was delineated above, the proposed systems for realizing overridable access restrictions always require rule-breaking to be *explicitly* acknowledged by the respective user and inform some enforcement instance of any single case of such an override having taken place. Even if the enforcement instance could still overlook some of these logs, this would significantly heighten the likelihood of rule-breaking behavior being detected, thereby increase the efficiency of motivation as compared to established enforcement practices, and ultimately lower the residual losses having to be borne by the organization.[18] From the more abstract perspective of agency theory already stressed above, such mechanisms of overridable access restrictions can thus be said to counteract the asymmetry-problem of *hidden action*.

This would, however, still leave the problem of hidden information about the respective context unsolved which leads to the principal not being able to make founded judgments about the adequacy of overriding actually being used and which, in the end, results in certain residual losses because of the problem of moral hazard. If, for instance, a user employs overriding in order to access a certain public hotspot, and if this fact is actually reported to the centralized enforcement instance, the adequacy of the respective user behavior can nonetheless only hardly be evaluated and the user can, in principle, always justify his doing by some kind of "emergency requirements", ultimately resulting in losses for the organization.

To overcome this problem and to reduce motivation costs even further, the approach of overridable access restrictions could be combined with that of logging contextual aspects already discussed in section 10.1.2. As outlined there, logging of contextual information can be employed to reduce asymmetries of hidden information and to allow for better ex-post evaluation of the adequacy of member behavior. If, for example, it is not only logged *that* a certain member has used the override mechanism to access a hotspot in an insecure manner but also *where* this happened or *what* the user was

[18] Note, however, the importance of non-repudiation of such logs in this context (see Ferreira et al. 2006, p. 850).

doing while being insecurely connected, then any kind of ex-post evaluation would better be able to decide whether overriding was reasonable in that context or not.[19]

Besides this reduction of asymmetries, overridable access restrictions would also counteract the problem of possibly ambiguous scopes of formal (and informal) rules. Different from these, part-time members or consultants would always be informed about the (in)admissability of their doings: Either they are granted access or not or they are informed about an overridable restriction through the respective warning having to be acknowledged explicitly. Similarly, overridable access restrictions would eliminate the need for rule-internalization to a certain extent that would otherwise lead to short-time members possibly not being aware of a certain rule at all. Even if not being completely self-enforcing, overridable access restriction thus ensure that any user is always informed about the currently effective rules and can be assumed to always do what he does consciously.[20]

On the other hand, the same drawbacks already discussed for context-enabled architectural means in section 10.1.1 would presumably apply here, too: For applying overridable architectural means within *really* interwoven environments, any single "external" contractor and any externally sourced service would have to be equipped with the respective technical solutions. Furthermore, "external" providers would presumably be unlikely to reveal extensive information about their inner workings and even if they would be willing to to so, evaluating the adequacy of individual behavior would be more complicated in such contexts.[21]

Another possible drawback regards the costs of actually exerting ex-post reviews of conscious rule-breaking. Different from traditional architectural means, overridable access controls are not self-enforcing but rather need a certain amount of explicit enforcement efforts having to be carried out. Even if these costs of exerting ultimate enforcement on a single case basis could be significantly decreased through the mechanisms described herein (as compared to traditional enforcement methods of formal rules), these costs have to be taken into consideration, too. In this regard, it should also be noted that there is some risk of making too much use of the "third mode" introduced by overridable mechanisms that could lead to overrides being used too often. This would then possibly result in the enforcement instance receiving too many notifications and, ultimately, lead to inappropriate judgments on the adequacy of some logged behavior.[22]

[19] Like for the case of context-logging, a multitude of further contextual aspects could be logged. Think, for instance, of repeated webcam photos documenting the difference between an isolated hotel room and a hotel lobby. Furthermore, logging of how long it took for the user to establish a secure connection after insecurely accessing a hotspot would also allow for better ex-post judgments. The latter would, however, again refer to the problem of hidden *action*.

[20] See, for example, Ferreira et al. (2006, p. 849): *"[The] alert message [makes the user] aware that he is trying to access information he is not authorized to see. This makes him responsible for what he is doing [...]."* See also Padayachee and Eloff (2007, p. 78), also highlighting that the user has to *"accept responsibilty"* for overriding an access restriction.

[21] Nonetheless, Pretschner, Massacci, and Hilty (2007) already proposed some first ideas regarding the use of comparable technical approaches to be used in SOA-environments.

[22] Ferreira et al. (2006, p. 851), for instance, mention this risk of managers receiving *"too many notifications"* of explicit overrides that could lead to situations where managers *"do not even bother to check"* any single notification.

Overall, we can thus state that overridable access restrictions could represent a less costly alternative for enforcing formal rules. They could be employed to counteract the high motivation costs that would arise from "traditional" enforcement of these rules because of the significant ex-post information asymmetries having to be expected for the fourth era. Explict ex-post inspections and sanctioning would still have to be realized and would result in some costs having to be borne, but these costs would be significantly lower than for the traditional case because the enforcement instance would be informed about any single case of an override having taken place and because of ex-post inspections only having to be made for those cases. Like the other rather hierarchical approaches of motivation discussed in sections 10.1.1 and 10.1.2, however, the approach of overridable access restrictions can hardly be applied to realize motivation within *really* interwoven environments of the fourth era that make extensive use of outsourcing, SOA or SaaS. But for the problems arising from increased mobility, context-dependence and ever-changing membership statuses, explicitly overridable access restrictions could represent a significantly less costly alternative to traditional enforcement approaches for formal rules.

10.1.3.3 Implications

Based on these considerations, we can now analyze the overall cooperation costs that would presumably arise from the approach of explicitly overridable access controls. As elaborated above, the approach makes strong use of delegating ultimate decisions to the individual members who are much more familiar with the specific givens and requirements of a certain situation. Coordination costs are therefore comparably low but still slightly higher than for the currently established mixed approach because of coordination decisions having to be made with explicit regard to one of the three available modes and thus leading to certain misallocations.

Motivation costs, in turn, could be reduced significantly by using this alternative approach for enforcing formal rules. Instead of having to make a costly trade-off between the probability of rule-breaking behavior being detected and residual losses arising from undetected rule-breaking behavior, organizations could reach – at least for certain scenarios – a high (if not nearly perfect) detection rate even under the conditions of strong information asymmetries. The risks arising from moral hazard could possibly be counteracted through context-logging in case of a restriction being consciously overridden. Even if the respective technical meachanisms would – like architectural means – require a certain amount of initial fixed costs to be borne, they could thus significantly lower the costs of ultimate enforcement and the residual losses as compared to traditionally enforced formal rules while at the same time preserving their expedient property of violability.

Nonetheless, the approach of overridable access restrictions will presumably not be applicable to *all* scenarios constituting the fourth era. In particular, aspects like SOA, SaaS or the increasingly important outsourcing arrangements will only hardly be addressable. In this respect, overridable access restrictions do not differ much from the other hierarchical approaches discussed in the previous sections. Regarding aspects like an increased member mobility, context-dependence and ever-changing membership

statuses, however, explicitly overridable technical means could very well counteract the significant increase of cooperation costs and serve as more efficient alternative to the now established mechanisms. Again, this might render the respective fourth-era practices valuable for organizations (see figure 10.4).

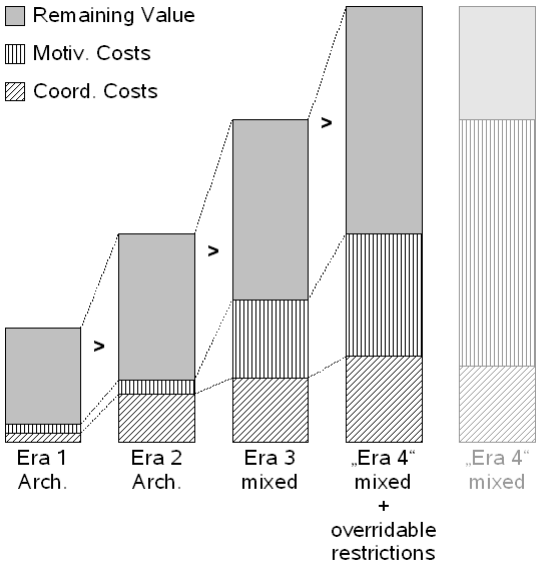


Figure 10.4: Possible cost advantage of explicitly overridable access restrictions over established approach.

There are, however, some further aspects that have to be noted with regard to this approach. First of all, logging plays an important role for the detection of possibly unfavorable user behavior and for the ex-post evaluation of the adequateness. Only through these logging mechanisms can the serious information asymmetries be overcome and the overall costs be reduced within this approach. Like it was the case for extensive and context-enabled logging discussed in section 10.1.2, however, it is questionable to what extent such logging practices would be compatible with the currently existing regulatory framework. On the one hand, the logs would contain a lot of information that could affect the users' privacy rights, but on the other hand, these logs would – different from the previous approach – only be made after the respective user had explicitly chosen to make use of the possibility to override a restriction and this affirmation could also be combined with some information that logging will be activated. After all, the model assumes that users are not forced to permanently make use of their overriding capabilities but that they can rather conduct most of their work without explicit overrides.

This, however, leads us to another aspect that will have to be discussed in the future: If coordination is to a reasonable extent delegated to individual members and if the collectively optimal behavior (which could only be identified by these individual

members) requires to violate a certain rule (be it through explicitly using an override mechanism or through simple disobedience of a "traditional" formal or informal rule) and if, finally, any such violation implies a certain individual risk of being prosecuted afterwards to the respective member, why, then, should this individual member violate the rule at all?

Of course, he could have an individual interest in doing so – saving time, for example – but whenever this is not the case, a rational user could be expected to refrain from (collectively desirable) rule-breaking and to play safe instead. This would be a completely different kind of selfish opportunism than that one usually assumed in the field of information security, but it could nonetheless lead to unfavorable outcomes and thus had to be counteracted, too. Even if neither this problem nor the possible solution approaches[23] shall be discussed in detail here, future work might also concentrate on this issue.

If, however, viable answers to this problem could be found, explicitly overridable access mechanisms might very well represent an efficient alternative for realizing cooperative member behavior. At least for some of the typical fourth-era scenarios, it allows to profit from delegation of decisions to individual members while at the same time raising significantly lower motivation costs than the nowadays established mechanisms for enforcing formal (and informal) rules.

Conclusion:
Overridable access controls can be understood as alternative mechanism for enforcing formal rules. Like the now established mechanisms, they would allow to profit from extensive delegation of ultimate decisions to the individual members but different from these, they would cause significantly lower costs of ultimate enforcement. Like the other approaches discussed in the previous sections, they can only be applied to some of the challenges characterizing the fourth era and like the approach of extensive logging, they might possibly conflict with the currently existing regulatory framework. If, however, these conflicts could be resolved, overridable access controls might represent an efficient means for realizing security-related cooperation.

10.1.4 Substituting Prices for Punishment

In section 7.1, we already mentioned Lessig's model of four modalities influencing human behavior. Three of these modalities (architecture, law and norms) were identified as analogons of our three meta-measures of architectural means, formal rules and informal rules. The fourth modality mentioned by Lessig, however, has so far not even rudimentarily been discussed as possible means for realizing motivation. This shall be caught up in the following.

Generally speaking, Lessig's fourth modality which he calls *"market"* influences individual behavior through monetary constraints that limit an individual's ability to

[23] Zhao and Johnson (2008, p. 6), for instance, suggest a bonus system to be used in this respect for rewarding those members that create collective value through conscious overriding.

engage in a certain behavior or, put more metaphorically, by attaching a price tag to the behavior that should be constrained. The prime example for the use of the "market" modality is the price of cigarettes that influences people's ability or willingness to smoke.[24] Like the "architecture" modality, this modality is said to regulate behavior *"more directly"* than laws and norms by Lessig (1998b, p. 664).

Throughout our previous considerations of motivational instruments, we already used the concept of a "price" being imposed on a certain behavior in a rather implicit and indirect manner. Remember, for instance, how we explained the way law – or, to use our meta-measures, a formal rule – influences individual behavior in section 7.1: We derived the "costs" of disobeying such a legal (or "law-like") regulation from the probability of rule-breaking behavior being actually recognized and sanctioned and from the severity of this sanction.[25] We then used the concept of an "individually perceived risk" of a would-be rule-breaker that affects his payoff-function and – as long as this risk supersedes his individual benefit from rule-breaking behavior – motivates him *not* to break the rule. At least indirectly, we have thus already used the concept of a "price tag" being consciously attached to "unwanted" behavior by a centralized instance for motivation purposes in a multitude of ways throughout the preceding sections.

What we have not done so far is using the instrument of prices *explicitly*. This does, however, not imply that this is impossible. In fact, such consciously imposed prices are already used in traditional economics in a multitude of cases that are – like our problem of security-related cooperation – subject to externalities. Pollution is the prime example from economic textbooks here: Instead of prohibiting a certain externality-causing activity (polluting the air) or limiting it to a well-defined, allowed amount, a tax is imposed on "any unit" of the respective behavior. A factory polluting the air (and thereby causing negative externalities to others), for example, has to pay a fixed price for any "unit" of pollution. Generally speaking, such taxes have two effects: First, they motivate the externality-causing party to reduce the negative external effect[26] and second, the taxes can be used to compensate those affected by the externalities.[27] Even if not being perfect for various reasons that will be examined in more detail below, such taxes are an established instrument for counteracting externality-related problems.

Now, remember that our problem of realizing security-related cooperation inside an organization also arises from the existence of externalities (see section 5.1).[28] All

[24] The price is, however, not the only mechanism through which the "market" constrains behavior. Quality, for instance, can do so, too (see Lessig 1999, p. 87). We will nonetheless concentrate on the *price* attached to a certain behavior herein.

[25] Remember that the "costs" resulting from these variables are not a simple mathematical product but rather subject to more sophisticated relations. See note 16 on page 156.

[26] The same principle could of course also be applied to positive externalities through subsidies. Let us nonetheless only consider negative externalities here.

[27] The abstract economic concept goes back to Alfred C. Pigou and is nowadays subsumed under the term of "Pigovian taxes".

[28] The basic similarity between the problems of information security and environmental pollution was, for instance, already recognized by Anderson (2002). Further mentions of the analogy include Camp and Wolfram (2000) or Anderson and Moore (2007, p. 2). This gives us additional

approaches discussed so far try to solve this externality-problem through mechanisms that use some kind of "price" indirectly being imposed on activities that were determined as disadvantageous from the collective point of view. The almost obvious possibility of levying such prices directly and explicitly has, however, not been considered so far. As a final hierarchical approach, let us therefore discuss the possibilities for doing so.

One approach for solving our problem on the basis of explicit prices would be a slightly modified version of the violable architectural means discussed in section 10.1.3 above. As mentioned there, this approach rests upon mechanisms like overridable access restrictions whereas any act of overriding has to be consciously approved by the respective user and whereas extensive logging and notifications to some centralized instance allow for ex-post evaluation of adequateness and – in case of the override being rated inappropriate – punishment.[29] Any case of conscious override is thus associated with a certain "indirect" or "virtual price" that is "paid" by the user and that results from the probability of being punished and the severity of the expectable sanction.

To change this approach of "virtual prices" being "paid" ex-post into one of "real" (tax-like) prices being paid ex-ante, think of an overridable architectural means that only allows overriding after the user has agreed to pay a certain amount of some currency that represents – analogously to pollution taxes – some kind of an "insecurity tax". This "insecurity tax" could then be withdrawn from an individual account of that user which could, in case of not being exhausted, be turned into an individual benefit. Analogously to the approach of taxing pollution, this would provide an incentive to engage less in the externality-causing activities but would at the same time allow the user to produce externalities whenever the derivable value supersedes the tax.

Such an approach seems at least worth to be discussed with regard to the coordination and motivation costs that would arise from it. Assume, therefore, a modified version of the violable architectural means where any conscious override requires the user to accept, instead of the measures aimed at ex-post enforcement mentioned above, a certain tax-like payment to be automatically withdrawn from some kind of "insecurity account". What would be the coordination and the motivation costs of such an approach?

10.1.4.1 Coordination

The most important challenge of such taxation schemes for externality-causing activities regards the determination of the optimal height of the tax. If the tax level is too low – when the tax is lower than the actual severity of the caused externality – it does not provide enough incentives for abstaining from externality-causing activities and when the tax is too high – that is, when the tax supersedes the impact of the externality – even those activities are prevented that would actually represent a benefit from

motivation for trying to transfer the taxation-approach to our problem of security-related cooperation.

[29] Note, however, that we assumed a "third mode" of access rights besides "allowed" and "prohibited" for these overridable architectural means. Of course, there will also be behaviors that simply should not be subject to overridable access restrictions. This is the case here, too.

the collective point of view. The optimal tax level would therefore exactly represent the overall impact of the respective externality.

Calculating this overall impact of an externality and setting the tax accordingly is in the original case of pollution to be realized by the state. For our case of organization-internal information security, this would then require some centralized instance to determine the height of the "insecurity tax" that had to be paid for, say, "accessing a hotspot in an insecure manner (for x minutes)". For such an "insecurity tax" to lead to preferable overall outcomes, this tax had, in turn, to be set to a level that represents the overall height of the resulting externalities as accurately as possible. Only then will individual members have an interest in abstaining from insecure behavior when the negative externalities supersede the derivable value while at the same time actually behaving insecure whenever doing so represents an overall benefit.[30]

This necessity for setting the "tax" to a level that represents the severity of externalities then leads us to a critical change that had to be made to the established principle of pollution-taxes in order to make it applicable to our problem of security-related behavior: While it is in the classical example of pollution comparably irrelevant for the severity of the external effect who pollutes, where he does so and under what circumstances, such factors are – as repeatedly outlined above – of critical relevance for the severity of negative effects within the field of information security. The risk being imposed to other members from a certain member (A) connecting to a hotspot does, for instance, not only depend on the question whether A connects to a hotspot in an insecure manner or not or on the duration of an insecure connection but also on a multitude of further factors like the number of additional active WLAN devices in the same environment, for example. Different from the traditional case of pollution, the severity of externalities being caused by a certain behavior thus strongly depends on a multitude of situational factors which, consequently, must have effect on the height of the "insecurity tax" being levied.[31] If explicit prices should represent the actual negative impact of "insecure" behavior, these prices have thus to be determined on the basis of some kind of a "situational risk factor".

Let us now return to our question regarding the coordination costs that had to be borne by an organization for the case of realizing information security by means of an "insecurity tax" as described herein. The coordination task having to be solved would in this case consist in determining the amount of "taxes" having to be paid for a certain behavior in a given situation. Even if some kind of "context-aware tax-calculation scheme" could be implemented into a technical solution that ensures enforcement, this would obviously raise the same problems of complexity, non-predictability and of ever-changing givens having to be incorporated in the coordination process that were

[30] Again, there also had to be some individual incentive that motivates members to *actually* spend for insecurity – as opposed to saving for individual profit – when this would be preferable from the collective perspective. Let us therefore assume that there is some kind of profit-sharing agreement in place. As we only want to discuss possible approaches from a generalized perspective, the detailed properties of these incentive schemes don't matter here. Generally speaking, however, incentive- and "taxing-" scheme had to be aligned with each other.

[31] Note, however, that existing pollution-taxes are also diversified to a certain extent – based on the chemical substance being emitted, for example. Note furthermore that some kinds of pollution are prohibited instead of being taxed.

already mentioned for the approach of context-aware architectural means in section 10.1.1.2. Like for these, this necessarily had to result in substantial coordination costs having to be expected which, in turn, could again be limited by means of generalization to a certain extent.[32]

There is, however, one advantage that could be expected for a scheme of "insecurity taxation" as compared to the rather "simple" context-aware architectural means from section 10.1.1: Even if the risks arising from certain behaviors under certain conditions had to be estimated ex-ante in both cases and even if this had to be done on the basis of a certain set of (generalized) anticipated situational givens, the approaches differ with regard to the evaluation of *benefits*. In the case of context-aware architectural means, these benefits also had to be anticipated *before* the respective situations emerge in order to make a context-dependent yes / no decision that is to be represented within the access control mechanism, for example. In the approach discussed here, only the risks had to be evaluated ex-ante, while the evaluation of benefits is shifted to the situation where the possibility for generating value actually appears. It is then the individual member who has to decide whether the derivable benefit exceeds the anticipated risk (which is represented as a price). As outlined repeatedly, this member will usually be able to make significantly better estimations in this respect (as he is much more familiar with the situational givens, including unanticipated aspects) and consequently, this will presumably result in less maladaption and thus in lower coordination costs than the approach of context-enabled architectural means. Generally speaking, the coordination costs arising from "insecurity taxation" can therefore be assumed to lie somewhere between those assumed for context-enabled architectural means in section 10.1.1.2 and those predicted for overridable architectural means in section 10.1.3.1.

On the other hand, the approach of "insecurity taxes" would also be subject to the same basic limitations as context-enabled architectural means. Service-oriented architectures, really interwoven organizational structures, constantly changing and often unknown instances having to be taken into account – all these aspects will only hardly be addressable in an adequate manner within the approach of levying "insecurity taxes" from individual "members". But whenever uncertainty about cost-benefit relation of a certain behavior primarily arises for the "benefits" part while the respective risks are comparably easy to estimate in advance – because of only few possible contexts having to be considered or because of the relevant contexts being easily generalizable, for example – the approach of "insecurity taxes" might prove highly valuable from the perspective of coordination costs.

10.1.4.2 Motivation

The motivation costs of this approach can, in general, be compared to those arising for context-enabled and overridable architectural means (see sections 10.1.1.1 and 10.1.3.2). Applying the basically architectural means of motivation to a multitude of temporarily associated "external" parties or externally sourced services might turn out to be too costly or even impossible. On the other hand, a system based on "insecurity

[32] See again figure 10.1 on page 236.

taxes" would make most (if not all) activities of ex-post evaluation and punishments dispensable, thereby reducing motivation costs as compared to the violable architectural means from section 10.1.3.

Furthermore, there might be a need for establishing the "currency" that is to be used for "tax payments". Think, in this respect, of some kind of "insecurity credits" that are assigned to the individual members and that must, as outlined above, represent an individual (monetary) value for the case of not being spent in order to actually motivate members to behave securely. Instead of having to exert punishments to individual members in case of inadequate insecure behavior, the organization thus had to distribute rewards to those members that actually behave in a secure manner. Again, this might represent a certain amount of motivation costs having to be borne by the organization. But alternatively, one could also think of arrangements where these "taxes" are "withdrawn" from the virtual account of a team or profit center that determines the amount of bonuses being paid, for example. These "taxes" could then be "transferred" to the respective virtual accounts of those teams or profit centers that are harmed by the negative externality. In this case, motivation costs would not appear in the form of additional payments having to be made – existing payments would rather be redistributed. The overall motivation costs of the approach of "insecurity taxes" would thus mainly depend on the arrangement being found for "payment". In general, however, they can be assumed to be on a level comparable to that of violable architectural means.

10.1.4.3 Variation: Tradable "Insecurity Credits" as Factor of Production

From the delineated model of "insecurity taxes", it would be just a small step toward what could be called "tradable insecurity permits". Again, the model is already practiced in the field of pollution and describes arrangements where any party is allowed to engage in a certain amount of pollution. This allowance is then represented by a virtual "pollution permit" which can be traded among the individual players.[33] Assume, for instance, a player A that could derive a certain benefit (x) from engaging in a polluting activity which is allowed by a permit held by him. Assume furthermore another player B that has already exhausted all his pollution permits but that could derive a benefit of y from being allowed to cause an additional unit of pollution. If the possible benefit for B now supersedes the value derivable by A (that is, when $y > x$) B could pay A an amount somewhere between x and y to buy the permit. Again, both parties would be better off and additionally, pollution would happen exactly where its value is greatest.[34]

Of course, this approach can also be transferred to our subject of security-related behavior. Assume therefore that some centralized instance does, instead of raising an "insecurity tax", assign a certain amount of "insecurity permits" to any individual member. Like in the above example of "insecurity taxes", the organization could then

[33] The approach traces back to Coase (1960).

[34] Again, any economic textbook should describe the mechanism and it therefore needs not to be discussed in more detail here. See, for instance, Mankiw and Taylor (2006, pp. 200 ff), Milgrom and Roberts (1992, pp. 304 f), or Furubotn and Richter (2005, pp. 115 f).

set some precalculated, context-dependent price for engaging in a certain activity under certain circumstances. This "price" would be expressed in the "currency" of "insecurity permits". Connecting to a WLAN-hotspot at a given hotel (and under further given contextual aspects) would then, for example, "cost" an amount of 10 "insecurity permits".[35]

Here, at the latest, we reached a point where we have reinterpreted our original problem of security-related cooperation from one that regards the production of an organization-internal *"public good"* (see section 4.1.2) into one that concerns *insecurity* as a *"private good"* which is consumed by the members of an organization as a classical factor of production and which needs to be allocated in the most efficient manner.[36] Again, this opens a wide field for future discussions and alternative research approaches that shall not be examined here. For our concern, it is enough to recognize the reinterpretation and the need for allocating this "private good" in an efficient manner.

This allocation, in turn, could be realized by making permits tradable. For instance, one could think of some kind of electronic marketplace. If, then, one member (A) could derive a benefit of 1 from 10 of his remaining "insecurity permits" and another member (B) could derive a benefit of 5 from the same 10 permits but has none left, A could sell its 10 permits to B for a (monetary) price of, say, 3 and ultimately, both parties would again be better off. Insecure behavior would then take place exactly where the ratio between the (anticipated and precalculated) risk and the derivable benefits is most advantageous for the organization.

Like in the case of an "insecurity tax", this would require the coordination instance to estimate the risks (the strength of the externality) arising from a certain activity under specific circumstances *in advance* and with all the inefficiencies already mentioned above. Even if it would, unlike in the "tax"-case, not have to set the ultimate (monetary) price, it had at least to determine *relative* prices, stating that WLAN usage at airport X costs twice as many "permits" as WLAN usage in hotel Y, for example. Furthermore, it had to determine the overall amount of "permits" that are emitted.[37]

The main difference between the "insecurity tax" and tradable "insecurity permits" thus lies in the fact that in the case of "taxation", the coordination instance has to determine the optimal (monetary) price for behaving insecure while in the case of "tradable permits", the coordination instance only has to set prices in a relative manner but at the same time must determine the overall amount of existing permits and thus,

[35] Comparable ideas of tradable permits for insecurity have, even if in another context, also been discussed in brief by Camp and Wolfram (2000), by Thomas and Amon (2007, p. 2) or by Anderson, Böhme, Clayton, and Moore (2008, p. 52). Both do, however, equal insecurity and pollution and do not recognize a necessity for situation-dependent weighting.

[36] See, in this respect, again Coase (1960): *"[T]he right to do something which has a harmful effect [...] is also a factor of production."*

[37] For reasons that will be outlined in section 10.2.1 in more detail, the question how these permits should be distributed among the individual members is a secondary one. However, the initial distribution could very well matter for various reasons. Coase (1960) therefore suggests an initial distribution of rights that represents some forecast outcome of (hypothetical) costless market transactions. Mapped to our example, this would suggest to emit more initial permits to salespeople than to members from R&D departments, for example.

the maximum amount of insecure behavior that should be present in the organization. Both approaches have their pros and cons which shall not be discussed in detail here.[38] Both approaches, however, could also be applied to the field of organization-internal information security.

10.1.4.4 Implications

As outlined above, the cost structures of the possible approaches described herein are comparable to those of overridable architectural means described in section 10.1.3. First of all, the coordination instance has to determine when the "third mode" should be used at all. Furthermore, and different from the approaches discussed so far, the coordination instance has to determine the (either absolute or relative) price that has to be paid for engaging in a specific activity under specific circumstances. More practical, however, seems the possibility to specify a function by which some technical mechanism then calculates the respective price. In any case, this price will – due to the well-known problem of efficiency losses – be subject to certain inefficiencies. These then represent additional coordination costs which have to be added to those arising for overridable architectural means. On the other hand, the overall coordination costs will presumably be lower than those of context-enabled architectural means as discussed in section 10.1.1 as the benefits-part of risk-benefit considerations does not have to be estimated ex-ante but is not considered until the possibility for generating benefit actually arises.

Motivation costs, in turn, could be significantly reduced as compared to violable architectural means. As explicit prices influence behavior on an ex-ante basis and due to the technical enforcement of payments assumed herein, there is no or only little need for ex-post inspections – at least with regard to the motivation of users to behave in conformance with the coordination outcome. Even if "insecurity taxes" as well as "tradable insecurity permits" are subject to the same restrictions of applicability as all other rather hierarchical approaches discussed so far, they could therefore represent another possibility for realizing security-related cooperation (see figure 10.5).

Most of all approaches presented herein, "insecurity taxes" and especially "tradable insecurity permits" would represent *hybrid* arrangements in the sense of section 3.3, combining elements from hierarchical and market-based cooperation to reach a higher overall efficiency. Especially in cases where it is comparably easy (or cheap) to estimate the risks arising from a certain activity in advance and where uncertainties primarily regard the derivable benefits, they could thereby have an advantage over the other rather hierarchical approaches discussed so far.

Besides these cost implications, one additional aspect should be mentioned in brief. Both approaches described herein, "insecurity taxes" as well as "tradable insecurity permits", represent an approach that employs *explicit* prices as motivational means instead of the implicit and indirect ones always considered at least for the motivational means of formal and informal rules so far. They thus use a completely different

[38] Taxes, for instance, are usually advantageous for externalities with a more or less well-known impact, while tradable permits better suit those situations where the impact is rather unknown and where the externality-causing activity should be limited to a given overall amount.

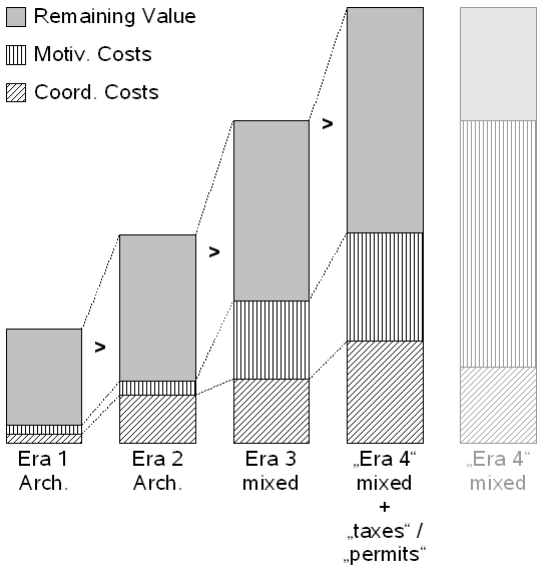


Figure 10.5: Possible cost advantage of “insecurity taxes” or “insecurity permits” over established approach.

approach of motivation than the currently established meta-measures. Furthermore, explicit prices are not only used for motivation purposes but also serve the task of *coordination* to a certain extent. At least far more than the other possible approaches discussed up to now, “insecurity taxation” and “tradable insecurity permits” thus rest upon the concept of an “invisible hand” determining individual behavior as compared to some “visible hand” of management that determines “right” and “wrong” in one way or another and that is more or less responsible for the ultimate outcome. With “insecurity taxes” or “tradable insecurity permits”, the ultimate outcome is therefore far less predictable and primarily determined by the play of the different forces acting within such a system. This approach would therefore go without *any* state of explicitly aspired member behavior that should be reached.

Such a fundamental change of approach would necessarily entail further questions to be answered and additional conflicts to be solved. In particular, the motivational means of explicit prices or of an organization-internal market for “insecurity permits” would in all likelihood not “fit into” the existing regulatory framework. This aspect will therefore be discussed in more detail in section 10.3. But these possible objections notwithstanding, the approach of “taxing” insecure behavior or of emitting “tradable insecurity permits” could very well be an additional possibility for counteracting some of the specific givens of the fourth era.

Conclusion:
"Insecurity taxes" or "tradable insecurity permits" represent a new meta-measure that uses explicit prices for motivation purposes. It allows to reduce motivation costs significantly while at the same time raising less coordination costs than context-enabled architectural means. Instead of anticipating the whole cost-benefit estimation, it only requires the coordination instance to determine the overall risk that would result from a certain behavior under certain situational givens. The task of estimating the derivable benefit, in turn, is carried out by the individual members. Where applicable, both variations of this approach might therefore represent viable ways for realizing security-related cooperation. Possible objections, however, regard the "compatibilty" with the current regulatory framework.

10.1.5 Hierarchical Approaches – Conclusion

Up to now, we have discussed four different approaches for counteracting the security-related implications of the fourth era as predicted in section 9.2.

- *Context-enabled architectural means* could allow to realize more aspects of motivation through technology again and thereby avoid the significant motivation costs of formal and informal rules otherwise having to be expected for the fourth era.
- *Extensive logging* of user behavior as well as of contextual aspects could reduce the strong information asymmetries that underlie these substantial motivation costs to a certain extent while at the same preserving the principle of delegation.
- The introduction of *overridable architectural means* can be interpreted as an alternative enforcement mechanism for the established formal rules that would also preserve delegation (and thus, the ability to take situation-specific aspects into account) but at the same time strongly reduce information asymmetries and thus motivation costs.
- And finally, the introduction of *"insecurity taxes"* or *"tradable insecurity permits"* would realize motivation on the basis of explicit prices, raise a certain amount of coordination but only few motivation costs and, most notably, would partially realize coordination on the basis of more market-like mechanisms. Both options therefore represent *hybrid* arrangements that combine rather hierarchical and rather non-hierarchical practices.

Of course, all of these mechanisms can be combined with each other in various ways. One example for such a combination is the suggestion to use context logging from section 10.1.2 together with overridable access restrictions in order to reduce ex-post asymmetries of "hidden information" given in section 10.1.3. Similarly, context-enabled architectural means could also be combined with the concept of overridability or with extensive context-logging, for example.

The options for realizing cooperation under the specific givens of the fourth era are multifold and the four examples given so far shall in no sense be understood as some kind of an exhaustive set. In fact, there might be a multitude of further approaches for realizing security-related cooperation under the specific givens of the fourth era in a hierarchical way.[39] Any such approach, however, could presumably be better assessed and compared with other ones after an economic examination like those pursued in the preceding sections. At least, this allows for better predictions about the expectable effectiveness and efficiency of the respective approach as well as it could reveal aspects that would possibly not be seen otherwise.

All approaches discussed so far, however, do not seem to be suited to the *really* profound changes coming along with the upcoming switch toward the fourth era. Even if aspects like increased member mobility and context-dependence as well as the ever-changing membership statuses could be counteracted to a certain extent through the above-mentioned mechanisms, none of these can actually be applied to settings of strongly interwoven, boundaryless "organizations" that make strong use of service-oriented architectures (including a multitude of "externally" obtained services), Software-as-a-Service, or extensive outsourcing.

This suggests the assumption that the basically hierarchical approaches discussed so far might simply be inappropriate for being applied to the significantly less hierarchical "organizational" structures predicted for the fourth era. We will not even try to prove or disprove this assumption. Instead, we will in the following discuss two less hierarchical approaches to give some suggestions how such rather decentralized cooperation mechanisms could look like.

Conclusion:
There exist various possible approaches for realizing security-related cooperation in a rather hierarchical manner under the conditions predicted for the fourth era. All approaches discussed herein, however, turned out to be unsuitable for the really profound developments toward actually interwoven, boundaryless "organizations". Less hierarchical approaches could possibly better meet the respective requirements and shall therefore be discussed, too.

10.2 Initial Thoughts for Less Hierarchical Approaches

In the preceding section, we discussed some mostly hierarchical approaches for realizing security-related cooperation under the specific givens of the fourth era. All of these approaches have in common that they do basically require some centralized instance which plays a more or less significant role for the overall cooperation process. But as we have seen so far, these approaches will only hardly allow to actually overcome

[39] Of course, there might also be reasons for applying some of these approaches within third-era settings, too. Overridable access restrictions, for example, could also prove valuable within traditional office environments.

the challenges arising from the most significant characteristics of the fourth era and to render the real era-switch economically reasonable for an organization.

There is, however, an alternative course of action for reducing the expected cooperation costs of the fourth era that shall be discussed in the following. Generally speaking, the initial ideas discussed below do, instead of trying to counteract fourth-era specifics, actually *accept* the characteristics of this era and support strong decentralization of the whole cooperation process in order to utilize situational, context-specific knowledge of the relevant individuals. Consistently with our economically inspired framework of security-related cooperation, this would require mechanisms for assuring that the coordination outcome actually conforms – at least as far as possible – to the collective interest of *all* "members" of the "organization". In particular, any mechanism has to solve the problem of externalities possibly arising from individual behavior that was outlined in section 4.1.1 and, with explicit reference to cooperation, in section 5.1. Furthermore, any mechanism also has to realize motivation of individual members to behave in accordance with the collective interest as determined during the coordination process.

Realizing cooperation through rather decentralized approaches implies that we can no longer assume some kind of centralized instance as primary authority for coordination and motivation. Instead, both processes will increasingly have to be realized among the individual members themselves in in a suitable way. From an abstract and strongly generalized perspective, we therefore do not have a certain number of principal-agent relations between some centralized instance and the individual members anymore but are rather more and more confronted with a multi-agent situation where any individual is principal and agent at the same time (see figure 10.6). At least from this strongly generalized perspective, the hierarchical, principal-agent-like approaches discussed so far seem unlikely to fit such settings. On the other hand, already established and rather decentralized approaches to information security (those usually associated with today's informal rules) could – for reasons outlined above – also not be used efficiently.

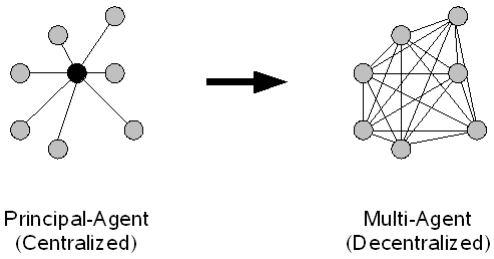


Figure 10.6: Principal-agent vs. multi-agent setting.

This does, however, in no way imply that there were no alternative ways for realizing cooperative behavior among multiple agents without some kind of centralized instance. In fact, there are, to speak with Denning and Hayes-Roth (2006, p. 23), *"plenty of alternatives to central decision making"* and again, most of them are well-known within

economics. In the following sections, we will therefore discuss two approaches that could be used for realizing security-related cooperation within the strongly decentralized, interwoven organizational structures of the fourth era.[40]

As outlined in chapter 3, cooperation among individuals can basically be realized through hierarchical arrangements or through market-like structures, where in the hierarchical case, individual behavior is guided by the *"visible hand"* of some centralized instance while market-like structures rest upon the *"invisible hand"* of the market mechanism and individual self-interest. Both of these cases, however, require some costs having to be borne in order to realize cooperation, which were distinguished as "hierarchy costs" and "market costs" in chapter 3.

As we have seen in the preceding sections, the costs of realizing security-related cooperation through hierarchical approaches will in all likelihood increase significantly under the emerging conditions of the fourth era. The two rather decentralized approaches discussed in the following therefore basically rest upon the principle of an *"invisible hand"* of some market-like mechanism being used for realizing security-related cooperation. For this reason, we will have to depart from some of the dimensions used so far to explain the cost structures of hierarchical cooperation. In particular, we will not further regard hierarchical efficiency losses but rather consider "market costs" as introduced in chapter 3. Furthermore, motivation costs also have to be seen in a slightly different light as the market mechanism itself already provides a certain motivational effect.

Even more than the rather hierarchical approaches discussed so far, the following approaches only represent first ideas and general, fundamental reflections and shall *not* be understood as propositions for practices that could already be applied to real scenarios. Due to this non-concrete nature of our considerations, we will also abstain from making statements about relative cost structures – as compared to established approaches or to the rather hierarchical approaches from above, for example – as such statements would hardly be credible within the rather fundamental considerations that will follow. Even if the discussed approaches might thus at first sight seem "far-fetched" or "inapplicable" to some readers, their discussion could at least open our mind for reflecting on possible approaches for realizing information security under the specific conditions of the fourth era.

10.2.1 Internalization of Externalities Through Mutual Bargaining

As outlined in section 5, the need for realizing cooperation among the different "members" of an "organization" with regard to security-related behavior basically arises from the existence of externalities. Due to these externalities, be they understood in a positive or in a negative manner (see section 4.1.1), the behavior of any single member affects the individual outcome for any other member as well as it influences the overall collective outcome. Generally speaking, the aim of any possible cooperation mechanism is to solve this externality problem. Economic theory, in turn, knows various

[40] Some of the ideas discussed in this section were originally presented in 2008 (see Glaser and Pallas 2008). In these cases, the discussion will, however, be more detailed herein.

ways for solving this problem in a more market-like manner instead of the basically hierarchical approaches considered so far.

The classical example from economic textbooks in this regard is that usually associated with the *"Coase Theorem"*: As Coase (1960) has shown, the problem of externalities could, in a hypothetical world without transaction costs, be solved through mutual agreements of economic exchange alone. Assume, for illustrative reasons, two players of which one (player A) exerts a negative externality[41] upon the other (player B). Assume furthermore that A derives a certain benefit from the activity that causes the (negative) external effect for B. If, under these givens, the negative effect for B supersedes the positive effect that A could derive from the respective activity, B could simply pay an amount lower than his injury but higher than A's benefit to make A abstain from the respective operation. In this case, the mutual agreement would result in the optimal overall outcome of A not engaging in the externality-causing activity. If, however, the benefit of A exceeded the harm resulting for B, B would not be willing to pay A an amount high enough to make him abstain from the activity. In the end, no agreement would be made between A and B and again, the optimal overall outcome would result: A would actually engage in the disturbing activity. The most astonishing aspect of this mechanism is that it requires *no* specific initial distribution of rights (for exerting externalities to others, for example) but only that *any* well-specified initial state is present. It is, for our example, absolutely irrelevant whether A has a right to engage in the disturbing activity or B has a right for not being disturbed. As long as there is any clear initial status, the outcome is always optimal within this theoretical model.

How, then, could such a practice of compensation be applied to our problem of realizing information security among a certain number of members? Assume, in a first step, an "organization" consisting of two freelancers (A and B) who want to engage in some cooperative activity and who use classical market-contracts and the "invisible hand" for realizing this cooperation. Assume furthermore that there is some well-defined initial state – perhaps codified in the contract – that forbids any insecure access to WLAN hotspots. Now consider a situation where A could derive a certain benefit from using a WLAN-hotspot because of significant time-savings. Let us say that this benefit is 10. On the other hand, A himself would also face a certain individual risk of being attacked via the unprotected connection. Let us assume that this risk has a value of 5. Finally, using the hotspot in an insecure manner would also cause an externality (a risk) to B which might represent costs of 1.

Even if there is an initial state prohibiting the insecure access to WLAN hotspots, it would, given this structure of values and risks, be more efficient from a collective point of view that A actually accessed the hotspot in an insecure manner instead of abstaining from its use: The overall value would be 10 while the overall costs (in the form of risks) would be $5 + 1 = 6$. Within the idealized model mentioned above, this situation could be reached if A and B agreed on a contract that allows A to access the hotspots but that at the same time obligates him to pay an amount above 1 (and

[41] Again, it does in this respect not matter whether the externality is a positive or a negative one, the mechanism works for both. We assume a negative externality for reasons of simplicity alone.

below 5, his individual value) to B. If, for reasons of simplicity, A and B agreed on a compensation of 3, both were better off as both had an overall advantage of 2 as compared to the initial "default" state (see table 10.1). Other values for costs and risks might of course lead to different efficient outcomes. The initial state (insecure access allowed or not), however, would *not* affect this outcome.[42]

Table 10.1: Simple compensation scheme – Two players

Overall Payoff	Player A	Player B
Initial state (no access)	0	0
Uncompensated access	$10 - 5 = 5$	-1
Access with compensation of 3	$10 - 5 - 3 = 2$	$-1 + 3 = 2$

Now recall the example of a ten-member "organization" already given in section 5.1. Within the same model and given the same values for risks and benefits, A now had to compensate any single of his cooperators for being allowed to access the hotspot in an insecure manner. As such an insecure access would again result in a negative external effect of 1 for *nine* individuals, overall compensation had to be larger than 9. In this situation, A would not be willing to pay such a compensation because his individual benefit from WLAN access would be smaller than the required compensation. A would therefore have to abstain from WLAN usage and again, the efficient outcome of A not using the hotspot would appear. If, however, A could derive a benefit of, say, 20, he could very well afford to compensate his nine co-members as his individual benefit $(20 - 5 = 15)$ would be larger than the cumulated individual risks of the co-members $(9 \cdot 1 = 9)$. If A paid a compensation of $1.5 \cdot 9 = 13.5$, he himself could still reach an individual payoff of 1.5 and again, this would represent the optimal outcome from a collective point of view.

All that sounds too god to be actually true. And in fact, the model given by Coase is still a theoretical one. In particular, it assumes an unrealistic world without any transaction costs and works perfectly only within this unrealistic world. But in the real world, such transaction costs of market transactions do very well exist in the form of "market costs" and additional "agency costs" which shall therefore be discussed in brief.

10.2.1.1 Coordination

As outlined in section 3.1 market costs refer to all costs arising from the use of the price mechanism and include, in particular, the costs for carrying out negotiations and the costs of closing a contract. Now, reconsider the above case of a ten-member "organization" of freelancers with one of these members (A) trying to access a hotspot in order to generate an individual advantage of 20. For being allowed to do so, he had

[42] In the real world, however, the initial state does very well matter in various respects. See, for instance, Furubotn and Richter (2005, pp. 104 ff).

to enter into individual negotiations with any of the other nine members. And any of these nine members had to calculate his individual risk that would arise from *A* accessing the hotspot in order to estimate his minimum compensation. Obviously, this calculation would either turn out to be very costly for the individual members[43] or the calculation outcome would in many cases strongly differ from the actual risk. The latter would, in turn, result in negotiations being made on the basis of wrong price-calculations and thereby presumably lead to strongly inefficient outcomes while the former would represent rather traditional negotiation costs which could also prevent individuals from reaching an efficient outcome through market transactions. And of course, the costs for realizing the necessary communication (including the time being spent) also had to be considered as costs of using the price mechanism for realizing cooperation.

Economic textbooks also highlight the fact of "Coasian" internalization being of at least limited applicability in a multitude of situations. In particular, this is the case when the costs for realizing an agreement exceed the overall benefit that could be realized. This, in turn, can especially be the case when the number of involved parties is high[44] or when the different individuals (have to) speculate about the value being ascribed to a certain outcome by the other individuals and bargain on the basis of these speculations.[45] Both of these characteristics apply to our case of WLAN usage within a ten-member-"organization" and will, in all likelihood, also apply to a multitude of real-world cases where security-related cooperation has to be realized. Furthermore, the usual cost-increasing effect of complexity will also be of strong relevance for the individual members' ability to calculate their minimum compensation.

This cost increase would presumably arise from the other phenomena of the fourth era, too. Within strongly interwoven "organizations" that make extensive use of external services, SaaS and outsourcing, the number of relevant entities possibly suffering (negative or positive) externalities from one individual's behavior would be certainly high and reaching an efficient outcome on the basis of mutual agreements would presumably become virtually impossible. Even the increasing relevance of context within otherwise unchanged organizations could hardly be addressed as the strength of a negative effect having to be borne by one individual because of a certain behavior of another – the induced risk – largely depends on these situational factors and is thus highly volatile and costly to estimate.[46]

Ultimately, we could therefore assume extensive market costs to arise from the theo-

[43] For instance, any single co-member had to estimate the probability of a negative externality arising for himself from *A*'s insecure WLAN usage in a specific situation whereas a crowded fairground would, for example, presumably imply a different probability of adverse events than, say, an isolated hotel in an industrial area. Furthermore, any single co-member also had to estimate the severity of the respective negative externality in order to calculate his individual risk that had to be compensated.

[44] See, for instance, Mankiw and Taylor (2006, p. 197) or Milgrom and Roberts (1992, pp. 302 f).

[45] This could, for instance, be done in order to take larger shares of the overall value generated through the bargain. See, for example, Furubotn and Richter (2005, pp. 105 f) or again Mankiw and Taylor (2006, p. 197).

[46] This resembles, in a way, the problem of determining the height of "insecurity taxes" and the "situational risk factor" discussed in section 10.1.4.1. We will come back to this issue below.

retical approach of coordinating security-related behavior through traditional, market-like exchange transactions among individual "members" under the givens of the fourth era. So far, this would clearly raise strong doubts on the efficiency and thus on the applicability of the bargaining approach for realizing security-related cooperation. We could then stop our considerations of the bargaining approach here.

But – and this is the point where we necessarily have to become rather speculative – if we could find a way to significantly reduce the mentioned market costs, this would possibly allow us to use mutual compensation agreements, anyhow. In particular, it was repeatedly mentioned throughout this work that technological progress can affect transaction costs in a multitude of ways and thereby influences the way cooperation is realized among different individuals. If this is the case, then there might also be a way to do so for those market costs that would otherwise prevent an efficient application of the bargaining approach described herein. And if we could identify such instruments, then we might very well be able to apply this approach to the field of security-related cooperation and to outline a possible alternative to the hierarchical approaches discussed so far. Let us therefore make some short reflections about the requirements that had to be fulfilled by a technology that should actually render the bargaining approach efficiently applicable.

First of all, let us consider the traditional costs associated with the explicit closing of a contract: communication costs, travel costs or the time spent in order to sign the contract are usually mentioned in this respect. Technology, however is able to reduce all these costs significantly. Communication costs can be lowered and contracts can be transmitted and signed electronically. In particular, it is also possible to delegate the whole bargaining process as well as the closing of (standardized) contracts to specialized software agents and to thereby eliminate large portions of these transaction costs in some areas. Especially in the context of Ubiquitous Computing, such (semi-)autonomous agents are already discussed intensively. Adopting this idea of specialized (semi-)autonomous agents for closing compensation agreements on security-related aspects could possibly represent a way to eliminate the "traditional" market costs within the field discussed here, too. If, for instance, any member were equipped with a software-agent that demands a certain amount of compensation from any member wanting to perform a certain externality-causing activity, and if these agents could negotiate with the demanding member's agent on some kind of electronic marketplace, then the compensation arrangement could be reached by these agents at virtually nonexistent costs.[47]

Another impediment for the application of mutual bargaining mentioned above regards problems arising in situations with many different parties being involved. If, for instance, one member had to close agreements with nine other ones on the respective compensations having to be paid for carrying out a certain activity, this might lead to deadlocks that prevent such agreements from actually being made. But if bar-

[47] Yes, the example of WLAN usage would seem somewhat misplaced here as the agents – as well as individuals – would need some communication channel to carry out such negotiations. But for other cases regarding SOA, for instance, this objection would not apply. And, not to forget, we are still *speculating* about fundamental approaches for realizing security-related cooperation in a decentralized manner, not proposing well-elaborated practices.

gaining were realized through software agents acting on some electronic market-place, such situations could also be avoided by proper protocols and bargaining schemes, for example. Additionally, such protocols might also allow to overcome the problem of speculation.[48]

What remains is the problem that a member who is affected by the behavior of another one is unable to estimate the strength of the externality. While in the traditional examples given by Coase – burned wood because of trains throwing sparks or crops destroyed by straying cattle – the severity of an externality can easily be estimated and thus be represented by a price, this is not the case for our problem of information security. If A wants to access a WLAN hotspot in a given situation and B should state some price for allowing A to do so, B must be able to evaluate his individual risk that would arise from A accessing the hotspot. This price then had – like any other risk – to be the result of some likelihood of an adverse event actually occurring and of the loss that would be caused when this event actually occurred. As mentioned above, both of these factors could only be estimated by B at certainly high costs.

If, however, bargaining activities were delegated to software-agents, these estimations also had to be delegated to them. A possible approach could then lie in the request of A's agent carrying extensive context information along with it in a trusted manner.[49] The bargaining request would then not only ask the other members' agents to specify a price for allowing "insecure access to a hotspot" but rather for allowing "insecure access to a hotspot a geolocation *xyz* at time *abc* while n other active WLAN devices were detected in the same area", for example. One could then think about bargaining agents that could calculate an individual risk for any single affected member on the basis of this information and of a multitude of further aspects. Based on this information, the bargaining agents could then ultimately come to an individual amount of required compensation.

Closer inspection of this idea reveals that the task having to be solved by such agents would require an *almost perfect* risk calculation to be realized on the basis of vast amounts of contextual information. In a sense, any single user agent had to solve the task of determining the height of taxes and, in particular, a "situational risk factor"[50] that was assigned to the centralized coordination instance for the approach of taxing insecurity in section 10.1.4.1. As outlined there as well as for context-enabled architectural means in section 10.1.1.2, any such calculation scheme will have to be based on a certain amount of generalization which, in turn leads to some contextual aspects being omitted from the determination of actual compensations.

If, however, relevant situational givens could easily be anticipated and generalized for being represented in some calculation scheme that is then implemented into the in-

[48] Elaborating these mechanisms in detail would unquestionably go far beyond the scope of this work and lead to a multitude of further considerations having to be exemplified. In particular, this would include the whole area of mechanism design in electronic systems. Those readers wanting to go deeper into this field might want to start with Nisan (2007).

[49] See, in this respect, the need for trusted logs addressed for ex-post evaluation on the basis of context-logging in section 10.1.2 above.

[50] Note that this "situational risk factor" had to be determined for *other* members by the respective software agent. If, for instance, A wants to access a hotspot, B's agent had to determine the risk-factor that results from A's current situational givens.

dividual members' software agents and if some kind of an electronic marketplace could be established where these software agents could interact with each other, then the approach of compensation bargaining might actually be used to realize coordination of security-related behavior without a centralized instance. The approach would then be comparable to that delineated in section 10.1.4 with the only difference that the explicit prices would not be set by some centralized coordination instance but would rather emerge from background bargaining processes among individual member-agents.

Of course, all this is highly speculative and somewhat diffuse but if the approach of mutual compensation-bargaining should be realizable at all, it would presumably have to be implemented along these lines.

10.2.1.2 Additional Motivation

Even if motivation is usually assumed to be largely realized by individual self-interest within a paradigm of market-based cooperation, additional motivation costs can nonetheless still arise. We will not discuss this aspect in detail here but some brief notes on motivation shall, for the sake of completeness, nonetheless be given.

If we assume that the whole coordination process is realized by software-agents bargaining over externality-compensations on some kind of electronic marketplace, we still have to ensure that individual "members" actually behave in accordance with the respective coordination outcome. In particular, a member whose agent did not agree to the compensations required by his cooperators' agents had to be prevented from nonetheless carrying out the respective activity. Furthermore, it had to be ensured that agreed compensations are actually paid. Basically, both aspects could, like any other act of motivation discussed so far, be realized through ex-post or through ex-ante motivation. For reasons of simplicity, let us nonetheless assume that there is some technical system that enforces the outcome of the bargaining process.[51]

First of all, any such technical system forcing one member (A) to behave in accordance with the "market"-outcome must be trusted by all other parties possibly affected by A's behavior. If this were not the case, A could simply engage in damage-causing activities without caring about compensation at all and the whole cooperation mechanism would be basically worthless. Furthermore, all would-be cooperators had to be equipped with such technical solutions – for realizing the bargaining process itself as well as for being trusted by the other parties – and all these solutions at least had to be interoperable with each other.[52] Any attempt to actually realize cooperation in the compensation-based way discussed here will have to take these basic requirements into consideration. Of course, there is again a multitude of motivational factors – the usual dilemmas known from agency and game theory, for example – that could be discussed here, but doing so would again go too far. Future considerations of comparable

[51] Of course, all the more or less sophisticated motivational measures and combinations of these that were already discussed above could basically be applied here, too. As, however, the state that is to be enforced can be assumed as being well-defined and as requiring no situation-specific interpretations, architectural means seem suitable in this respect.

[52] One could, for instance, think of the various possibilities for using Trusted Computing technologies in this respect. This would, however, again open a whole new field for future research that shall not be entered herein.

approaches, however, will also have to take such factors into account.

10.2.1.3 Implications

There is a multitude of further aspects that could be discussed with regard to the approach of realizing information security through compensation bargaining. But due to the (currently) highly speculative and diffuse nature of this idea, we will leave it at the state of reflection reached so far: Basically, bargaining over compensations for externality-causing activities could realize coordination in a decentralized manner and reach efficient outcomes. This would, however, require market costs to be nonexistent or at least sufficiently low. Future technologies in the form of bargaining agents might be applied here and allow for a sufficient reduction of transaction costs. Such agents also had to be able to evaluate the individual risk that would arise for "themselves" in case of another entity engaging in some externality-raising activity. This would, in turn, require the agent to have some trusted knowledge about the contextual givens under which the respective activity should be carried out.

Due to an inherent need for generalization and because of the impossibility to anticipate any aspect that could possibly become relevant, however, a certain amount of maladaption costs still had to be expected. These costs might render the approach inappropriate for various cases, but for settings where context is less relevant or easily generalizable and where centralized cooperation mechanisms could – for whatever reason – not be applied, compensation bargaining might very well represent a viable alternative for realizing security-related cooperation.

In any case, however, the approach of compensation bargaining might provide a rich source for future ideas with regard to security-related cooperation in decentralized settings.

Conclusion:
Mutual compensation agreements between different players represent a decentralized alternative to centralized mechanisms for solving the problem of externalities. This approach could, with technological support of software-agents, possibly be applied to security-related externalities, too. Doing so would, however, require the context to be less relevant or easily generalizable. If this can be assumed, compensation bargaining might represent a viable alternative for situations where hierarchical cooperation can, for whatever reason, not be applied.

10.2.2 Agency Theory Revisited: Screening, Signaling and Individual Liability

Besides mutual bargaining over compensations, one could also think of other approaches for realizing security-related cooperation under the specific givens of the fourth era. As discussed above, compensation bargaining would require individual parties (or their bargaining-agents, respectively) to be able to assess the negative impact that would presumably arise for themselves from security-related activities of

others. This would, in turn, require vast amounts of trusted contextual information to be known and to be taken into account during the calculation of necessary compensations. Ultimately, this necessity led us to the above conclusion that relevant contextual givens had to be easily foreseeable and generalizable for compensation bargaining to represent a viable decentralized alternative to the hierarchical approaches discussed in section 10.1.

If, however, these contextual givens cannot be anticipated and generalized, this would in all likelihood render the approach of compensation bargaining impractical and for these situations, we therefore had to find other ways of realizing security-related cooperation in a decentralized way. In this section, we will thus outline some abstract ideas about possible starting points for the future development of alternative decentralized approaches that might also be applied to those settings that prove incompatible with the need for anticipation and generalization. Even more than it was already the case for the approach of compensation bargaining, these ideas only represent incipient stages and shall again in no way be understood as directly applicable practices.

For the development of these basic approaches, we begin with some strategies that are already established in the field of outsourcing, one of the areas that were repeatedly identified throughout section 10.1 as not being adequately addressable by hierarchical approaches to security-related cooperation. Starting from these approaches, we can then speculate about possible ways to adopt them to the problem of realizing security-related cooperation among different individuals in a decentralized manner.

As we will see, most of the security-related practices already established in the field of outsourcing have well-understood representations in classical agency theory. Think, therefore, of our "externals", part-time "members" or even of all "members" as very small but independent economic entities providing outsourcing services to each other through bilateral contracts. This would be consistent with our perception of a multi-agent setting as outlined above.[53] Any single entity would in this model be principal as well as agent to a multitude of other entities and all these entities still had to solve the tasks of coordination and motivation in order to cooperate with each other. How, then, is this realized in the field of outsourcing and how could the respective approaches be adopted to other fourth-era settings?

10.2.2.1 Coordination

Up to now, we always understood coordination as the problem of determining how a given individual member "should" actually behave with regard to information security from a collective point of view. But alternatively, and especially with regard to the decentralized, fourth-era settings discussed herein, one could also understand the task of coordination as asking which agent should be chosen for cooperation from a variety of diversely behaving ones. The task of coordination would then be to assign functions to the "right" cooperators. This is exactly how coordination is realized by the "invisible hand" of the market mechanism and how this task is usually solved in the field of outsourcing.

[53] See figure 10.6 on page 264 and accompanying text.

Take, for instance, activities of due diligence, which are usually taken out before choosing an outsourcing provider to cooperate with. These activities are aimed at the identification of the outsourcing providers' "quality" which means, for the field of information security, the identification of the security level the diverse providers are at least able to provide. Within classical agency theory, this resembles the concept of *screening*. Like due diligence, screening is carried out by the principal before delegating tasks to an agent in order to assess the agent's "quality" and to thereby estimate the possible quality of the outcome of the agent's activities. Regarding the various information asymmetries within principal-agent relations distinguished in section 3.2.2.1, screening (and thus activities of due diligence, too) is thus aimed at the reduction of *hidden characteristics*.

Besides screening activities, agency theory suggests these hidden characteristics to be counteracted through the instrument of signaling. Basically, signaling rests upon quality signals that are sent out by the agent and that allow the principal to rate the agent's quality without extensive inspection. For such signals to be a meaningful basis for quality ratings, it has to be less expensive for "good" agents to emit signals of high quality than for "bad" ones.[54] Only then can signals actually be used by a principal to limit the problem of hidden characteristics.

Within the field of outsourcing (and other forms of inter-organizational cooperation) and with regard to information security, such "meaningful signals" are, for example, realized on the basis of established information security standards or, more precisely, on the basis of certifications against these standards. An organization that has received such a certification does, by presenting it, emit a signal about its security-related qualities. As long as it is more costly for an "insecure" outsourcing provider to obtain such a certification than it is for a "secure" one, a principal can justifiably assume the certified provider to be more able to behave in a secure manner than the non-certified one. Even if such signals do not necessarily allow for a well-founded estimation of the absolute level of information security being realizable by the certified party, they do at least enable the principal to make a comparative rating of different providers and thereby limit the problem of hidden characteristics even further.

Both abstract concepts, screening as well as signaling, are thus well-established in the field of outsourcing. Let us therefore consider how they could possibly be adopted within the other settings predicted for the fourth era.

For the concept of screening, this would require individual players to be able to assess another entity's ability to provide a service in a way that imposes a low level of negative externalities on the client.[55] Based on this evaluation, the player with the best abilities for averting negative externalities could then be chosen for cooperation. This would, however, give rise to the problems of complexity, information asymmetries and context-dependence repeatedly discussed above: A high complexity together with strong information asymmetries among the different members will presumably lead to screening outcomes that do not represent the actual abilities of behaving secure (and thus, of preventing negative externalities) provided by the screened entity and thereby

[54] See Spence (1973). Note, however, that this condition is a necessary and not a sufficient one.
[55] Again, the same assessment could also be made with regard to positive externalities, respectively.

at least limit the value of screening activities.

Context-dependence, in turn, invalidates the screening outcome even further: To predict the agent's capabilities for keeping the level of negative externalities low, the screening party had to anticipate a multitude of possible contexts and to predict the ability of the screened entity to behave securely in any of these situations. As *sole* mechanism for choosing the agent to cooperate with, screening thus seems to be inappropriate. Nonetheless, screening might still prove valuable as one mechanism among others for comparing different possible cooperators against each other with regard to their security-related capabilities.

Signaling, in turn, would necessitate the existence of meaningful signals which are trusted by all entities and which are actually obtainable at reasonable costs by those agents that are able to cause less negative security-related externalities than others. While this role is in traditional outsourcing agreements filled by established information security standards, these would hardly be applicable to settings where cooperation is to be realized among individuals or other small economic entities.[56] The substantial costs having to be borne for such established certifications would presumably not pay off for such players even if their ability to behave in a secure manner is high and the level of expectable negative externalities could thus, in principle, be low as compared to other players.[57]

More appropriate would therefore be less comprehensive signals that nonetheless allow to assess the signal-emitting agent's capabilities of externality-prevention. Attestations of having attended security-related trainings and passed accompanying exams would be a simple example for such "lower-level" signals that could also be used by individuals or small groups to signal their security-related "quality". Another possibility would be the establishment of reputation mechanisms through which an agent could signal his capabilities to produce few externalities on the basis of former cooperators' experience with him. And finally, one could also think about the establishment of more lightweight information security standards and certifications that are more suitable to the small entities predicted for the fourth era. As the currently established, rather all-encompassing frameworks will in all likelihood prove inappropriate for signaling purposes under the predicted givens of the fourth era, a possible path for future developments would thus be the identification or creation of such signaling mechanisms that can also be used by smaller entities or even by individuals to inform possible cooperators about their security-related qualities.

With the exception of reputation mechanisms, all these approaches would, however, only allow to estimate a possible cooperator's *capabilities* to behave in a way that causes a low level of negative externalities to others. But neither screening nor signaling mechanisms provide reliable predictions about the *actual* behavior or even about the level of externalities having to be expected for the case of cooperating with a given agent. The mere capability of behaving secure does not necessarily imply that a would-be agent actually makes use of this capability. This is where mechanisms of motivation come into play.

[56] Recall, in this respect, the size-decreasing effect of technological progress described in section 3.4.
[57] See, in this respect, also Ghose and Rajan (2006).

10.2.2.2 Additional Motivation

Both approaches discussed so far are primarily aimed at the problem of *hidden characteristics* and support the principal in his selection of the agent to cooperate with. Even if this reduction of ex-ante information asymmetries already allows to reduce the losses arising within a model of strictly market-based cooperation, none of the above-described practices solves the various problems arising from *hidden action* and *hidden information*, which are primarily subject to the task of *motivation*. Even if the principal is able to select the agent with the highest "quality" – that is, for our topic, the agent with the highest ability to minimize negative externalities – this would still leave the problem of motivating the agent to actually use these capabilities unsolved.

This motivational problem can not only be addressed through explicit supervision of activities and contextual givens but also through incentives that align the individual interest of the agent with those of the principal. Such incentives can, in turn, again be realized in a multitude of ways of which only one shall be considered here: the approach of profit sharing and risk transfer.[58] Such agreements do, broadly speaking, assign a certain amount of the value generated by the agent to the agent himself. This motivates him to actually generate profit and to react to situation-specific givens in a way that heightens overall benefit. On the other hand, the agent also has to hold a certain amount of risks, implying that he has to bear the negative effect of adverse conditions. In the end, this limits the agent's possibilities for shirking and thus reduces the risk of moral hazard.[59]

Service level agreements (SLAs) are a well-established instrument within the field of outsourcing that can be interpreted as such an arrangement of profit sharing and risk transfer: If an outsourcing provider is, for instance, paid on the basis of the availability of the service provided by him, payment schemes can be designed in a way that motivates him to act more in the interest of the principal than it would be the case for fixed payments. On the other hand, such agreements also transfer the risk of unanticipated and adverse events to the outsourcing provider who, for instance, also gets lower payments when he fails to provide a service because of interrupted power supply. This does, in turn, motivate the outsourcing provider to take steps that minimize downtime and to install capacities for emergency power supply, for example. In this case, the outsourcing provider is fully accountable for the outcome of his doings, independent of the contextual givens under which he realized this outcome. The externalities possibly arising from the activities of the outsourcing provider are internalized and the interests of provider and client are aligned with each other.

Directly transferring the concept of service level agreements to the cases discussed here would, however, presumably turn out to be insufficient. As the concept rests upon ex-ante agreements over the height of payments that have to be made in dependence on some previously specified measurement parameters like service availability

[58] The whole field of such incentives cannot be addressed in detail within this work. See alone Laffont and Martimort (2001) for a first idea of what had to be considered in order to realize completeness in this respect.

[59] See, for instance, Laffont and Martimort (2001, pp. 145 ff) or Milgrom and Roberts (1992, pp. 206 ff). See also section 7.5 on page 178.

or response time, service-level agreements can only be applied to issues where such a clear measurement parameter actually exists.[60] This can unquestionably not be assumed for any case of different parties causing security-related externalities against each other.

Think, for instance, of hotspot usage again: A service level agreement between two cooperating parties that should internalize the negative externality arising from insecure WLAN access would, first, require to anticipate the possibility of one party actually using a hotspot at all and to recognize the fact of a negative externality possibly arising for the second party. Even if this anticipation could be assumed, the involved parties had, second, to agree upon some measurement parameter that relates the fact of insecure hotspot access to the severity of the resulting externality (the induced risk) and ultimately, to the height of payments or to some discount, respectively.

Obviously, we would then again be confronted with the problem of calculating a risk value in advance for a multitude of possible activities that could take place in a multitude of possible contexts. Alternatively, we could also think of some kind of calculation function for this value that cooperators could agree upon. Both options were already discussed in sections 10.1.4 and 10.2.1 and as shown there, both would presumably lead to considerable inefficiencies because of an inherent need for anticipation and generalization. If, however, the relevant factors could easily be anticipated, generalized and measured, SLA-like agreements might be a viable alternative to (or, more correct, a viable variation of) the approach of compensation bargaining supported by software-agents that was discussed in section 10.2.1. For other settings, however, we have to identify alternative ways for realizing the risk transfer.

The repeatedly discussed problems arising from ex-ante calculations or estimations of the height of an externality suggest to think about the possibility of determining the severity of a negative effect *ex-post*, after such an externality actually appeared. Instead of being based on rough ex-ante estimations, the amount having to be transferred from the externality-causing party to the party suffering from the externality would then be determined on the basis of the damage that was actually caused. This damage, in turn, might in many cases be easier to assess than the rather inappropriate ex-ante estimations of risk-values discussed so far.

Such an approach would clearly resemble the concept of *liability* already stressed for the hierarchical case in section 7.5: Any harm being imposed by one party on another had to be compensated with a payment that exactly represents the severity of the damage. Any externality would then be internalized and any party could be expected to act in the collective interest. For such a practice of liability to lead to optimal outcomes, however, three conditions had to be fulfilled: Any case of harm-causing had to lead to a corresponding (ex-post) payment actually being made, the height of a caused damage had to be exactly measurable, and we had again to assume costless transactions between the liable and the injured party.

For all three issues, real conditions differ significantly from these idealized assumptions and consequently, we can expect considerable costs to arise. In particular, detec-

[60] Recall in this context, the discussion on inadequate monitoring proxies from section 3.2.2.2.

tion of an externality will presumably be imperfect as well as the identification of the respective externality causer, measuring the impact of an external effect is likely to be rather an estimation than a "measurement", and due to the necessity of evaluating the height of caused damage and enforcing the respective compensation payments ex-post, we can expect a certain amount of costs here, too. All these aspects and the possibilities for reducing the respective costs could very well be subject to future research[61] but discussing them all in detail would again go too far for this work. Instead, let us consider one additional issue which regards the resulting planning activities having to be performed by the liable party:

Due to the obligation to compensate any caused externality ex-post, this party had to calculate some expected level of liability payments that would presumably arise from a certain activity (or non-activity) ex-ante in order to decide whether to engage in this activity or not. If, for instance, a liable individual "member" had to decide whether to access a hotspot in an insecure manner or not, this member had to be able to at least estimate the probability of harming others in case of hotspot usage and the height of compensation payments having to be expected for this case. This would, in turn, require him to be aware of all relevant aspects influencing these values. In particular, he had to know a multitude of facts that lie outside of his own sphere but rather characterize the systems used by his cooperators, for example. It makes, as a very simple example, a significant difference whether a cooperator has a firewall installed or not for the calculated probability of having to pay a compensation and thus for the decision of using a hotspot or not.

Like non-anticipatable and non-generalizable contexts were identified as impediments for determining a demanded compensation ex-ante in section 10.2.1, a similar problem would thus – in a somewhat reverted manner – arise for the case of liability-based motivation. Even if the individual player knows his own current situational givens quite well and could incorporate these givens in the process of balancing individual benefit against expected compensations, the local givens of his cooperators do again play an important role for this trade-off and these can, due to the strong information asymmetries assumed for the fourth era, again not be known by him.

Realizing security-related cooperation on the basis of compensations having to be paid ex-post would therefore seem practical only when these relevant local givens of other cooperators could – analogously to the ex-ante bargaining approach from above – be easily included in the local decision over carrying out a certain activity or not. But if these relevant local givens of potentially harmed parties are comparably constant, situation-independent and easily generalizable as compared to the local givens of the party potentially causing a negative externality, and if we could assume a high detection probability and a good measurability of caused harm, then the approach of individual liability and ex-post compensations might possibly prove advantageous over the approach of ex-ante compensation bargaining.

There are, however a multitude of further factors that had to be discussed in this

[61] Think, for instance, of adaptations from the fields of intrusion detection systems and (computer) forensics and, again, of extensive logging in order to heighten detection probability and the chance of identifying the respective externality causer. Another aspect would regard the question *when* a negative effect should actually be considered as externality caused by another party.

respect and that cannot be addressed in detail here. But as it was not our aim to present application-ready practices but rather to point out possible future directions in an abstract manner, we can close our initial considerations on the approach of liability here and leave these factors open to future research.

10.2.2.3 Implications

As we can see, agency theory can not only be applied to the case of security-related cooperation being realized in a hierarchical manner. Instead, it might also prove highly valuable for the development and the discussion of possible, rather decentralized approaches. The cooperation challenge that will presumably have to be solved during the fourth era can be reinterpreted as one where a multitude of individual players serve as outsourcing providers for each other, leading to any party being principal as well as agent to a multitude of other cooperating parties. With such an interpretation, the task of coordination can be understood as one of choosing the right agents to cooperate with. This task could, in turn be supported by means of screening and signaling of which screening was shown to presumably raise considerable costs. Signaling, in turn, appears to be a highly interesting subject for future considerations: If even the small entities that can be expected for the fourth era could emit established, commonly trusted "meaningful signals" regarding their security-related "quality", this could significantly lessen the costs of distinguishing "good" cooperators from "bad" ones and thereby foster the use of nonhierarchical models of coordination.

The discussion of motivational aspects, in turn, could also profit from existing knowledge from agency theory in a multitude of ways. Even if we only discussed the approach of risk transfer through individual liability in brief and left out other practices like the deposit of bonds, we could identify some preconditions for such schemes to be profitably applied to the fourth-era settings assumed herein. In particular, the agent deciding whether to carry out a certain, potentially externality-causing activity had to be able to easily evaluate those local givens of his cooperators (the potential "victims" of negative externalities) that are relevant for his probability of actually having to pay a compensation later and for the expectable height of such a possible compensation. If, however, these givens could easily be assessed, then the approach of individual liability might actually be used for realizing cooperation within a decentralized multi-agent setting that does not allow cooperation to be realized on the basis of ex-ante bargaining over compensations as outlined in section 10.2.1.

In any case, a multitude of further questions had to be left open within our short sketch. Most obvious are the above-mentioned, motivation-related issues of detection probability, of determining the height of damage caused by an external effect and having to be compensated, and the necessity of enforcing that compensations are actually paid. For ex-post, liability-like compensations to represent a viable way for realizing cooperation, these problems need to be resolved at reasonable costs. The likelihood of conflicts to arise between externality-causing and suffering party, for example, will presumably require some third instance to carry out ex-post deliberations about causality, compensation height and comparable aspects. Again, the involvement of such a third, conflict-resolving party will cause a certain amount of costs that had

to be subject to cost considerations for the approach of ex-post liability, too.

Less obvious but at least similarly important is the so far unregarded possibility of highly complex interrelations among a multitude of different players that might lead to less distinct givens of causality and accountability. If a negative effect only arises because of two services being combined in a specific manner while any of these services alone would raise no such negative externality, for instance, identifying a liable party for the negative effect might become impossible and in the end, the damage might stay uncompensated.

Even if the application of established agency theory to the decentralized multi-agent settings considered herein thus leads to an almost endless cascade of further questions and unsolved problems, it might nonetheless turn out to be valuable in a multitude of ways. An interesting task for the future would therefore be to map existing concepts from the field of information security to well-examined, abstract concepts from agency theory. Such a mapping might allow for a deeper, more abstract understanding of current practices of information security, illuminate possible conflicts that can be expected for future conditions or even suggest to modify current approaches in a specific way to meet the changed conditions having to be expected for the future.

Such a mapping between an established practice and abstract agency theory was herein made for the concept of information security standards and certifications, which were identified as one possible way for emitting "meaningful signals" with regard to security-related qualities of the emitting party. These meaningful signals are, in turn, of high importance for the selection of cooperators and thus for the coordination part of cooperation. Agency theory then allows us to make well-founded suggestions about requirements that will have to be met by such signaling mechanisms in order to support decentralized ways of security-related cooperation under the specific givens of the fourth era and ultimately suggests that currently established certification frameworks will in all likelihood not meet this requirement. These frameworks will therefore have to be complemented by less exhaustive, more lightweight certification schemes that can be used for meaningful signaling by smaller economic entities, too.

This, then, leads us to the very last aspect that shall be discussed in this work before proceeding to our summarizing conclusions: The implications of our deliberations for the currently established regulatory framework.

Conclusion:
The reconsideration of established agency theory might prove valuable for the development and discussion of future approaches to security-related cooperation in a multitude of ways. Practices like due diligence or standards and certifications have abstract and well-established representations in this field and can be examined on the basis of these abstract concepts, too. One implication of such considerations is that established certification schemes will presumably be of lower value under the specific givens of the fourth era and should be complemented with more lightweight alternatives. In any case, agency theory opens up a nearly endless field for conducting future, security-related research on the basis of a well-established economic theory.

10.3 Some Final Remarks on the Regulatory Framework

Throughout the preceding sections, we have, starting from our initial case of using public WLAN hotspots, identified a much broader development toward a fourth era of "interwoven computing" that is currently getting under way. Phenomena like the increasing relevance of mobility, outsourcing, temporary or multiple membership statuses or even technological developments toward service-oriented architectures and software-as-a-service are blurring the established organizational boundaries in a multitude of ways.

Based on this recognition of a fourth era, we elaborated some expectable implications for the realization of security-related cooperation under the changed circumstances. Following our economically inspired model developed in the previous chapters, we distinguished these implications into those regarding the task of coordination and those affecting established motivational measures. We then exposed why it would at first sight seem appropriate to make strong use of decision delegation in this fourth era (the specific givens would otherwise presumably lead to material maladaption costs) but at the same time identified strong reasons to believe that the established motivational approaches associated with such delegation (formal and informal rules) would lead to extensive motivation costs and thereby render the switch to the fourth era unreasonable, too.

This dilemma then motivated us to think about possible alternative approaches for realizing security-related cooperation under the specific givens of the fourth era. We therefore discussed various approaches on the basis of our model and discussed the overall cooperation costs that could be expected for the application of these approaches. Generally speaking, we found that all these alternative practices could reduce cooperation costs in one way or another as compared the hypothetical case of current practices being simply maintained. Any of the discussed approaches could thereby possibly render fourth-era practices economically reasonable at all and could thus prove advantageous for organizations.

This would, however, give rise to a conflict that has so far not been discussed in proper detail and explicitness: In section 2.2, we did not only identify our three initial meta-measures that play such a central role throughout the whole work, but also recognized the existence of a regulatory framework that limits an organization's choice of means for realizing information security. This regulatory framework consist of a multitude of legal conditions, quasi-legal standards and other restrictions and obligations that do in various ways define what an organization is allowed or even obliged to do with regard to information security. The regulatory framework thus defines the "field of the game" for realizing security-related cooperation within an organization. It defines some practices as obligatory and prohibits others.

Now reconsider the above-mentioned implications of the fourth era. As argued in section 9.2.2, the now established meta-measures of formal and informal rules will in all likelihood turn out to be highly inefficient under the changed conditions predicted for the fourth era. For an organization, it might therefore be reasonable *not* to realize security-related cooperation on the basis of these meta-measures but to choose one of the above-mentioned alternative approaches instead.

Even if doing so seems absolutely appropriate from the perspective of realizing cooperation in an efficient manner, the existing regulatory framework would, however, in many cases *forbid* the organization to actually take this course. An organization that has, for whatever reason, to prove compliance with the ISO standard 27001 would, for example, have to conduct awareness trainings or to draft a formal policy with regard to the use of mobile computing[62] in order to comply with the set of security controls prescribed by this standard – even if these controls turn out to be largely inefficient because of the changed conditions characterizing the fourth era.[63] Other regulations having to be followed might prescribe other inappropriate controls to be in place and ultimately, an organization might be forced to consciously waste resources for pursuing practices that are known to be strongly inefficient. Alternatively, organizations might even be hindered from the adoption of fourth-era practices at all because of a third-era regulatory framework prescribing third-era techniques. Both cases will hardly be desirable and thus call for careful reconsideration of the currently established regulatory framework.

Such careful reconsiderations are also indicated for another reason that accrues from the above discussion of possible alternative approaches. As any of these alternative approaches does not only go without some currently established and potentially prescribed practices but also introduces new ones which fit the specific requirements of the fourth era much better in one way or another, these newly introduced practices might also conflict with the established regulatory framework.

This is most obvious for the approach of extensive context logging discussed in section 10.1.2. As outlined there, such extensive context logging might significantly reduce information asymmetries between centralized instance and individual member and allow for better founded ex-post evaluations of the adequacy of member behavior. Extensive context logging might therefore counteract the otherwise considerable motivation costs of formal (and possibly of informal) rules and thereby make the adoption of fourth-era practices economically reasonable for an organization.

On the other hand, and as already stated in section 10.1.2.3, extensive logging of contextual information would under most legal regimes presumably violate the privacy rights of the respective member and thereby violate the regulatory framework which consists of privacy laws, too. Different from the above-mentioned case, the given regulatory framework would at this point not prescribe an inefficient and unreasonable practice but rather prohibit a practice that would otherwise allow for more efficient outcomes. This, then, leads to the question whether such reductions of efficiency are "acceptable" from a general, rather societal perspective or whether the given regulatory framework should be changed in order to allow for a higher efficiency to be realized.

While this question is – hopefully – comparably easy to answer for the case of extensive, detailed logging of user behavior and the respective contextual information, things might be different for the approach of violable architectural means discussed in

[62] See ISO / IEC (2005b), annex A, control 11.7.1 and 8.2.2. Indeed, the real world of auditing is slightly more sophisticated and rests upon even more extensive frameworks like CobiT (see ITGI - The IT Governance Institute 2007). The general conflict, however, remains the same.

[63] See, in this context, also Siponen (2006a), questioning the effectiveness of established practices with regard to security standards.

section 10.1.3: As the individual member can in this case decide on his own whether to override a restriction in a controlled manner or not, it is also left to the member's decision whether extensive logging actually takes place. From this point of view, it could be argued that there is no conflict with the existing regulatory framework.

But on the other hand, a member could also be induced to override a restriction and thus to allow context logging even if he actually disapproves the associated collection of substantial information. As mentioned in section 10.1.3.3, there must be a reason for the member to actually make use of his overriding capabilities (instead of playing safe) when doing so would presumably lead to an overall benefit for the organization. Let us therefore, without going more into detail, assume that there is some kind of a bonus scheme that associates conscious overrides with the possibility of obtaining an individual monetary benefit. In this case, the member would possibly have to decide whether to insist on his privacy rights or to "voluntarily" abandon these rights in exchange for a monetary benefit. We would then find ourselves right in the center of a highly moral and legal discussion that shall not be explicated in detail here. For our purpose, it is sufficient to recognize that the approach of overridable architectural means also holds the possibility that conflicts with the regulatory framework might arise.

Different than extensive context-logging and overridable architectural means, the approach of "insecurity taxes" or "tradable insecurity permits" discussed in section 10.1.4 represents a completely novel practice that is not even rudimentarily envisaged within the current regulatory framework and does therefore – metaphorically speaking – not "fit" into it (see figure 10.7). Nonetheless, explicit prices have been shown to represent a highly interesting alternative to established practices of coordination *and* motivation and might therefore seem worthwhile for various organizations. This would, however, again lead to possible conflicts with the current regulatory framework.

First, the introduction of such novel approaches would again suggest to abolish other practices that prove strongly inefficient under the changed circumstances. This would then lead to the same conflict of possibly having to maintain outdated practices without material reasons already discussed above. Besides this objection, the technical solutions suggested in sections 10.1.4.1 and 10.1.4.3 both require some information (the amount of "payments" having to be made, for example) to be transmitted to some centralized instance. This information might also hold implicit information about the respective member behavior and could thus – at least to a certain extent – be subject to privacy considerations. Besides these two types of possible conflicts, the approach of realizing security-related cooperation on the basis of *explicit* prices would presumably raise a multitude of further unsolved issues because of its general novelty.[64]

With the exception of context-enabled architectural means, all rather hierarchical alternatives would thus presumably lead to more or less serious conflicts with the currently established regulatory framework.[65] For some cases (like for extensive context

[64] For a first idea of the possible disputes, think – for a short moment – about an organization's attempt to convince an external auditor of the effectiveness of its "tax-based" approach to the use of public WLAN hotspots. The absence of some state of explicitly aspired member behavior that was brought up in section 10.1.4.4 would presumably be of certain relevance here.

[65] Context-enabled architectural means, however, represent an only slightly modified version of the

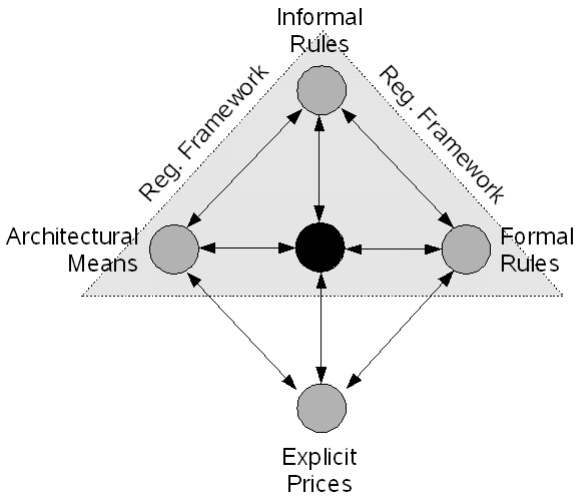


Figure 10.7: Explicit prices as motivational means not fitting into existing regulatory framework.

logging), these conflicts seem to be justified as they avert practices that would strongly affect the personal rights of the members. In other cases, however, the current regulatory framework might circumvent the application of practices that would otherwise prove highly valuable for organizations as they would allow to overcome the strong inefficiencies that will presumably arise from the ongoing shift toward the fourth era of interwoven computing.

Besides these rather hierarchical approaches, we also elaborated some initial ideas for realizing security-related cooperation in a less hierarchical, decentralized manner under the conditions of the fourth era. Even if these ideas have to be understood as early steps that would need much further consideration for being actually applicable, we could nonetheless derive some possible conflicts with the current regulatory framework here, too. First of all, we could for both approaches, ex-ante compensation bargaining as well as ex-post liability, expect the already mentioned problems that would arise from the abolishment of third-era practices. These would, however, in all likelihood emerge in a strongly exacerbated form as such new "organizational" structures would fundamentally contradict a multitude of the current regulatory framework's basic assumptions. Like the approach of explicit prices mentioned above, the discussed decentralized approaches would thus hardly "fit" into the current regulatory boundaries. On the other hand, such approaches could possibly turn out to be more efficient under some of the conditions predicted for the fourth era. In this case, it could again be indicated to think about possible modifications of the current regulatory framework to avert large-scale resource wasting.

already existing architectural means. But as already depicted in section 10.1.1.2, the possible fields of application are strongly limited for this approach.

Furthermore, decentralized, ex-ante compensation bargaining as drafted in section 10.2.1 would require any individual wanting to engage in a certain activity to communicate a multitude of contextual information to all his cooperators (or to their bargaining agents, respectively) all the time. This could again raise a multitude of conflicts with existing privacy regulations and therefore strongly limit the applicability of this approach. And finally, the application of abstract principles from agency theory does at least suggest to establish additional, more lightweight certifications that would – different from current information security standards, for example – allow smaller economic entities to emit meaningful signals about their security-related capabilities, too.

The basic ideas for rather decentralized practices discussed herein do, however, not even rudimentarily represent applicable approaches. Other strategies might be developed in the future and fit into the existing regulatory givens much better than those initially discussed above. The same is true for the rather hierarchical ones: Here, too, alternative approaches might emerge that are more compatible with existing regulations and nonetheless provide significant economic advantages over current, third-era practices. Our considerations were, however, in no way intended to represent an "exhaustive" set of possible fourth-era alternatives but should rather illustrate what alternative practices could be thought of in general and how the economically inspired model developed herein could be applied for the prospective evaluation of such alternative approaches.

But whatever approaches will be developed in the future for addressing the upcoming challenges: The switch from the third to the fourth era will only be possible with a regulatory framework that does not make the current status quo of information security practices a permanent one but that allows alternative, more appropriate practices to be used as well. Like the switch from the second to the third era was accompanied by a change from rather technical to rather management-driven information security standards and legislation, the upcoming era of interwoven computing will in all likelihood again require significant changes to be made to the regulatory framework. Such changes will, if being possible at all, take long time, need the involvement of a multitude of stakeholders and require extensive discussions to be led. An economic understanding of information security might, in turn, support and enrich these discussions in a multitude of ways.

But as the first of all steps, the need for a further development of the regulatory framework has to be recognized at all by the relevant players. The arguments developed herein might also provide some support in this respect. At least from an economic perspective, it seems virtually negligent to be satisfied with the current status quo of the regulatory framework.

Conclusion:
The different alternative approaches discussed herein would presumably conflict with the current regulatory framework in a multitude of ways. Due to these conflicts, the application of otherwise advantageous practices might be inhibited while the use of established but inefficient and unreasonable practices could be prescribed. This does at least indicate to think about possible adjustments of the current regulatory givens to better meet the changed requirements of the upcoming fourth era of interwoven computing.

Chapter 11

Conclusions

Chapter 11
Conclusions

One never notices what has been done; one can only see what remains to be done.

– Marie Curie

We have now arrived at the final chapter of this work. In the following, we shall therefore recapitulate our course of considerations and summarize the main contributions that we made to the field of organization-internal information security.

We began our considerations with the motivating initial case of using public WLAN hotspots in a professional context. As we have shown in chapter 1, most WLAN hotspots cannot be accessed in a way that complies with the well-established principle of strict network isolation. We discussed several possible approaches to this conflict and identified a lack of well-founded theoretical methods for choosing among these options.

We then noted that this lack does not only exist with regard to our specific case but rather for the field of organization-internal information security in general. We therefore decided to establish such an abstract, theory-funded understanding on our own and to do so by developing an economically inspired, positive model of information security inside organizations.

In part I, we laid out the theoretical foundations for this model. We developed two different approaches for structuring the field of information security and its current status quo: The scheme of different *"eras of information security"* as outlined in section 2.1 takes a historical perspective and sets prevailing computing paradigms and prevailing security practices in relation to each other. As a second scheme, we developed the *"security triangle"* in section 2.2. This security triangle distinguishes between the different meta-measures of architectural means, formal rules and informal rules and characterizes them by their enforcement model (ex-ante vs. ex-post) and their strictness. Additionally, it highlights the existence of a higher-level regulatory framework that restricts an organization's choice of means for realizing information security.

In chapter 3, in turn, we presented a consolidated view on those aspects of New Institutional Economics that are relevant for the development of our positive model of organization-internal information security: Transaction costs, information asymmetries, principal-agent relations, organizational structures and their various interdependencies were discussed in this respect. Furthermore, we did in this chapter introduce a

model of cooperation among individuals that basically distinguishes between the two tasks of *coordination* and *motivation*. Both partial problems have to be solved in order to realize cooperation but do at the same time also raise costs. These costs strongly depend on various factors like the number and complexity of coordination decisions having to be made, the level of uncertainty and information asymmetries or even the technological basis on which cooperation is realized.

Chapter 4 then presented some initial and already established economic aspects of information security that are of general relevance for the subsequent chapters. In particular, we noted the relevance of the economic concept of externalities for the field of information security, asserted that organization-internal information security shows characteristics of an organization-internal "public good" and distinguished four factors (direct and indirect costs and benefits) that determine the overall payoff resulting from a certain level of information security.

In part II, we then developed our economically inspired positive model of information security inside organizations. Based on the foundations established in the first part, we recognized that information security is basically a problem of economic cooperation among individuals and that this cooperation is today primarily realized by means of *hierarchical* cooperation (see chapter 5).

We therefore considered hierarchical coordination and motivation costs for realizing information security in more detail. In chapter 6, we combined the model of the three eras developed in section 2.1 with our remarks on hierarchical coordination costs from section 3.2.1 and came to an abstract understanding of how security-related coordination costs have developed over time. In particular, we were able to explain the observed shifts of prevailing security practices by the change of economic parameters (complexity, information asymmetries, etc.) that resulted from the repeated changes of prevailing computing paradigms.

In chapter 7, we then made a comparable mapping between the model of the security triangle from section 2.2 and the hierarchical motivation costs from section 3.2.2. We found that the different meta-measures have different costs structures and that these cost structures are differently affected by factors like the frequency of changed coordination outcomes or the level of information asymmetries. Additionally, we interpreted the use of ex-post-based motivational means (formal and informal rules) as an act of delegating ultimate coordination decisions to individual members.

The resulting economic model of organization-internal, security-related cooperation was then applied to the initial WLAN case for demonstrative purposes. This application led us to recommendations for addressing the problem of hotspot usage that coincide with the course of action that would presumably also have been chosen by intuition. Even if this could be interpreted as a verification for the adequacy and validity of our model, we would at this stage hardly have gained significant new insights of practical relevance.

Part III was therefore devoted to the *prospective* application of our model. In chapter 9, we identified various developments pointing toward an upcoming fourth era of *"interwoven"* computing. The increasing relevance of general member mobility, outsourcing, part-time and temporary membership and the technological movement toward concepts like service-oriented architectures (SOA) or software-as-a-service (SaaS)

constitute this new era. These developments will in all likelihood blur established organizational boundaries in a multitude of ways and thereby lead to significantly changed givens under which security-related cooperation will have to be realized in the future.

The expectable implications of these developments were outlined in section 9.3: As we have shown, the now-established meta-measures of formal and informal rules would lead to significant *motivation* costs under the changed conditions and would therefore presumably render the adoption of fourth-era practices inefficient. Established architectural means, however, would lead to substantial *coordination* costs and thereby prove inefficient, too.

Due to this expectable inappropriateness of established practices for realizing security-related cooperation, we proposed some alternative approaches in chapter 10 and discussed them on the basis of our economically inspired model. We showed that the different approaches counteract different factors of the expectable developments and that they will therefore prove appropriate for different kinds of challenges. Furthermore, we suggested on the basis of our model that none of the proposed hierarchical approaches will be efficiently applicable to situations with strongly interwoven, actually boundaryless "organizations". Such settings will presumably require less hierarchical approaches. Some initial thoughts on the development of such rather decentralized approaches were therefore presented, too.

Finally, we reconsidered the role of the regulatory framework. As shown in section 10.3, the currently established set of standards and legal rules would in various ways deter organizations from actually adopting fourth-era practices. In the end, this would lead to a deliberate use of inefficient approaches and thus to a conscious waste of resources. We therefore argued for a reconsideration of the current regulatory framework.

This logical structure is – for a last time – depicted in figure 11.1.

11.1 Main Contributions

After having recapitulated the logical structure of this work, we can identify some main contributions that we made to the field of organization-internal information security.

First of all, the development of an abstract, consistent and theory-founded understanding of information security inside organizations has to be mentioned in this respect. The application of established concepts and principles from New Institutional Economics leads to well-founded explanations for past changes of information security practices. It allows to make substantiated predictions about expectable implications of technological progress and provides a sound basis for formulating normative arguments.

Furthermore, the economically inspired, positive model developed herein associates the field of information security with knowledge about the general economics of organizations. It thereby illuminates the various interrelations between those areas and provides an integrated view to organization-internal information security. The economic understanding developed herein thereby provides new perspectives that can complement existing ones in a multitude of ways. This might not only support scien-

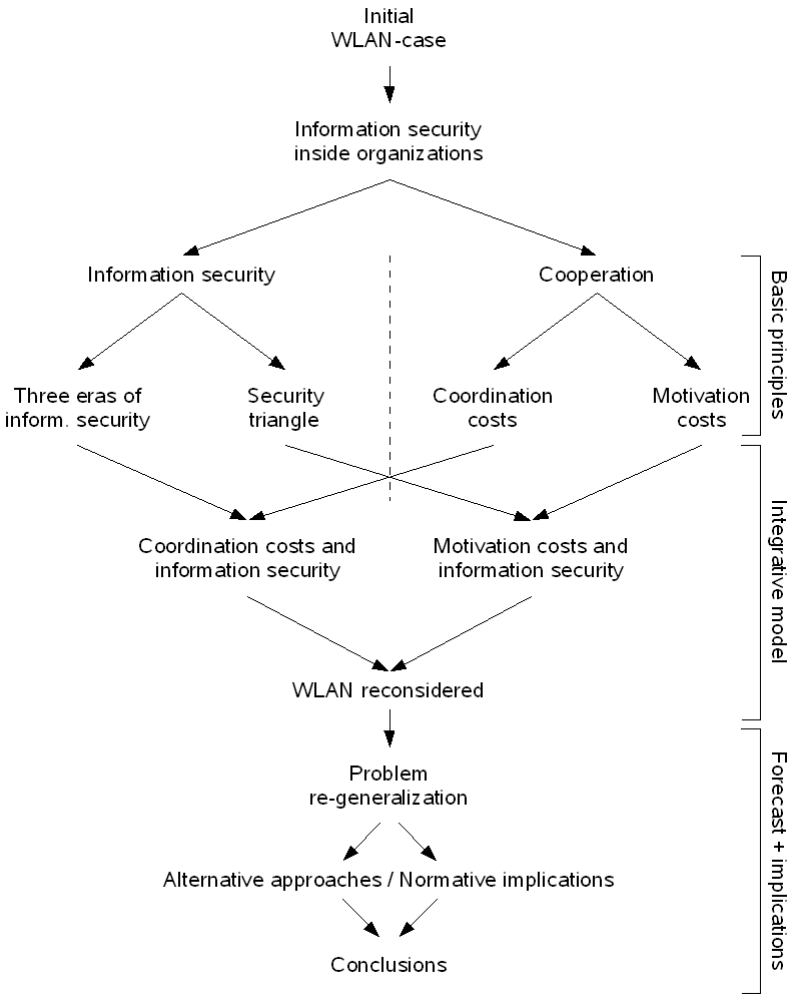


Figure 11.1: Logical structure of the work.

tific progress but could also lead to alternative approaches for teaching organization-internal information security in a more abstract, more consistent, but at the same time also more understandable manner.

In addition to this methodological contribution, this work also provides findings of more direct practical relevance. Most important in this regard is the prediction that the established motivational means of formal and informal rules will become strongly inefficient under the changed givens that can be expected for the future. Even if this finding might at first sight seem counterintuitive to most security practitioners, it results from well-founded theoretical considerations and will in all likelihood have consequences for future practices of "organization"-internal information security.

And finally, the discussion of possible alternative approaches might be of strong relevance for all those that are engaged in the development future strategies to in-

formation security and for those designing specialized technical systems in this field. Both parties could profit from the theoretical examination of the various approaches and derive implications for their own doing in a multitude of ways.

Besides these main contributions, this work provides many further aspects that can influence the academic as well as the practical world of organization-internal information security. In particular, there are various possibilities for conducting future research on the basis of this work. These possibilities shall also be outlined in brief.

11.2 Future Work

One starting point for future research could be the further consideration of established concepts from agency theory and their application to the field of information security. Signaling mechanisms, for example, were already discussed in brief in section 10.2.2. If it were possible to establish meaningful signals through which individuals and other small and mainly independent economic entities could communicate their ability to behave secure, this would presumably prove advantageous for the selection of members within traditional organizations as well as for the realization of security-related cooperation within fourth-era "organizations".

Bonding agreements, in turn, are an established instrument in classical principal-agent settings but have not yet been discussed in detail with regard to information security. Future work might however find ways for applying the concept of bonding agreements to security-related problems. And the perception of fourth-era practices as multi-agent settings that was introduced in brief in section 10.2 could also form a basis for future research. In any case, the application of agency theory to the field of organization-internal information security opens up a nearly endless field for conducting future research on the basis of well-established economic knowledge.

Another possible field for further research regards the internalization of formal and informal rules. This internalization was mentioned as a possible means of optimization in sections 7.2.2.1 and 7.2.3.1 but was at the same time stated as not yet being considered on a theoretical basis. Future research might be able to close this gap and provide deeper and theoretically founded insights into the possibilities for using formal and informal rules and for fostering their internalization. This could, for instance, be done on the basis of empirical studies on the (short- *and* long-term) effects of awareness-campaigns and security trainings. As outlined in section 7.5, such research could prove highly valuable for information security practitioners as well as it could lead to a better theoretical understanding of formal and informal rules.

And finally, we basically distinguished the two modes of hierarchy-based and market-based cooperation in in chapter 3 and made all further considerations on the basis of this dichotomy. These two modes of cooperation are often illustrated by the two metaphors of the *"visible hand"* of the manager and the *"invisible hand"* of the market. Nonetheless, many economic scholars are by now convinced of the existence of a third mode for realizing cooperation which is primarily based on trust among the cooperators. This third mode is usually illustrated as an *"invisible handshake"*.[1] An in-depth

[1] See, for instance, Ciborra (1985).

analysis of this third mode and its relevance for the realization of information security inside organizations could also be a highly interesting subject for future research. In this and in many further regards, the newly arising field of behavioral economics (see, for instance, Frey and Stutzer 2007) could also play a highly important role. The first "Interdisciplinary Workshop on Security and Human Behaviour" that was held in 2008 might represent a step into this direction.

In any case, a work like this can of course only represent a first step and has to leave a multitude of questions open and many aspects unconsidered. Future research can, however, profit from the economic understanding and the positive model developed herein in a multitude of ways.

Final Remarks

We opened this work with two quotations. The first one stated that *"the problems of IS/IT security management [are currently] tackled without drawing on existing theories"* and the second one warns against *"long-term messes"* that tend to *"become so familiar they look normal to most people living in them."*

The latter of these quotes characterizes the current world of organization-internal information security quite well. Security practitioners often find themselves in a messy world of standards, best practices, and rather intuitive "management" approaches. In most cases, they accept this mess as given. Instead of calling the mess into question, they change themselves and adapt to the messy world around them. This is, of course, absolutely understandable. In most cases, practitioners have no choice.

But this is only half the truth. Many scholars of information security management also seem to have accepted the mess and to build their own world around it. Instead of doing what scholars can do best – applying existing theories to novel problems and developing new theories – the scientific community largely approaches information security inside organizations by means that have their origin in the mess. And to make things worse, this is not only the case for research but also for the field of education. Some remarkable exceptions notwithstanding, theory-founded approaches to organization-internal information security do for the most part play a secondary or even nonexisting role in the academic world.

Nonetheless, it has been shown that the study of information security inside organizations is not necessarily confined to the application and development of standards, best practices and to other rather intuitive approaches. The whole field of information security inside organizations is very well open to scientific treatment that draws upon established and stable theories. Doing so requires a certain amount of curiosity, patience, creativity and the courage to get rid of old habits. But in the end, it might allow us to deal with issues of organization-internal information security in a more abstract, more consistent and, ultimately, in a more scientific manner.

Some might have accepted the existing mess. These might have concluded from the mere fact that information security management *is* tackled without drawing on existing theories that it *cannot* be done in other ways.

Fortunately, this it not the case.

Part IV

Appendix

List of Tables

List of Figures

Bibliography

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics 84* (3), 488–500.

Al-Muhtadi, J., A. Ranganathan, R. Campbell, and M. Mickunas (2003). Cerberus: A Context-aware Security Scheme for Smart Spaces. In: *Proceedings of the First IEEE International Conference on Pervasive Computing and Communications*, pp. 489–496. IEEE Computer Society.

Alberts, C. and A. Dorofee (2002). *Managing Information Security Risks: The Octave Approach*. Addison-Wesley.

Anderson, R. J. (2001). Why information security is hard – an economic perspective. In: *Proceedings of the 17th Annual Computer Security Applications Conference (ACSAC 2001)*, pp. 358–365. Online: `http://www.acsac.org/2001/papers/110.pdf` [18.05.2006].

Anderson, R. J. (2002). Unsettling Parallels Between Security and the Environment. The First Workshop on the Economics of Information Security (WEIS 2002). Online: `http://www2.sims.berkeley.edu/resources/affiliates/workshops/econsecurity/econws/37.txt` [23.02.2009].

Anderson, R. J. (2008). *Security Engineering: A Guide to Building Dependable Distributed Systems* (second ed.). Wiley.

Anderson, R. J., R. Böhme, R. Clayton, and T. Moore (2008). Security Economics and the Internal Market. Report for the European Network and Information Security Agency (ENISA). Online: `http://www.enisa.europa.eu/doc/pdf/report_sec_econ_&_int_mark_20080131.pdf` [10.03.2009].

Anderson, R. J. and T. Moore (2006). The Economics of Information Security. *Science 314* (5799), 610–613. DOI: 10.1126/science.1130992.

Anderson, R. J. and T. Moore (2007). The Economics of Information Security: A Survey and Open Questions. Conference on the Economics of the Software and Internet Industries, Toulouse. Online: `http://www.cl.cam.ac.uk/~rja14/Papers/toulouse-summary.pdf` [22.02.2007].

Anderson, R. J., F. Stajano, and J.-H. Lee (2001). Security Policies. Online: `http://www.cl.cam.ac.uk/~fms27/papers/2001-AndersonStaLee-policies.pdf` [14.09.2006].

Andress, A. (2004). *Surviving Security: How to Integrate People, Process, and Technology*. Boca Raton, London, New York, Washington DC: Auerbach Publications.

Andriessen, J. H. E. and M. Vartiainen (Eds.) (2005). *Mobile Virtual Work: A New Paradigm?* New York: Springer.

Ardagna, C. A., M. Cremonini, E. Damiani, S. D. C. di Vimercati, and P. Samarati (2006). Supporting Location-Based Conditions in Access Control Policies. In: *Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security*, Taipei, Taiwan, pp. 212–222. ACM. DOI: 10.1145/1128817.1128850.

Arrow, K. (1969). The Organization of Economic Activity: Issue pertinent to the choice of market versus nonmarket allocations. In: *The Analysis and Evaluation of Public Expenditure: The PPB System.*, pp. 59–73. Washington DC: US Government Printing Office.

Aviram, A. (2006). Network Responses to Network Threats. In: M. F. Grady and F. Parisi (Eds.), *The Law and Economics of Cybersecurity*, pp. 143–192. New York: Cambridge University Press.

Bailey, D. (1993). Managing Complexity in Secure Networks. In: *Proceedings on the 1992-1993 Workshop on New Security Paradigms*, Little Compton, pp. 2–6. ACM. DOI: 10.1145/283751.283762.

Balachandran, A., G. M. Voelker, and P. Bahl (2003). Wireless Hotspots: Current Challenges and Future Directions. In: *Proceedings of the 1st ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH '03)*, San Diego, pp. 1–9. ACM Press. DOI: 10.1145/941326.941328.

Ball, C. J. (1971). Communications and the Minicomputer. *IEEE Computer 4*(5), 13–21.

Becker, G. S. (1978). *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press.

Becker, G. S. (1993). Nobel Lecture: The Economic Way of Looking at Behavior. *The Journal of Political Economy 101*(3), 385–409. Online: `http://www.jstor.org/stable/2138769` [28.08.2008].

Bejtlich, R. (2007). No ROI? No Problem. Online: `http://taosecurity.blogspot.com/2007/07/no-roi-no-problem.html` [23.08.2007].

Bell, D. E. and L. La Padula (1976). Secure Computer System: Unified Exposition and Multics Interpretation. Published by ESD/AFSC, Hanscom AFB, Bedford. Online: `http://csrc.nist.gov/publications/history/bell76.pdf` [31.08.2006].

Bellovin, S. M. (2006). On the Brittleness of Software and the Infeasibility of Security Metrics. *IEEE Security & Privacy 4*(4), 96. DOI: 10.1109/MSP.2006.101.

Benson, D. H. (1983). A Field Study of End User Computing: Findings and Issues. *MIS Quarterly 7*(4), 35–45. Online: http://www.jstor.org/stable/248745 [10.07.2008].

Bharghavan, V. and C. Ramamoorthy (1995). Security issues in mobile communications. In: *Proceedings of the Second International Symposium on Autonomous Decentralized Systems*, pp. 19–24. DOI: 10.1109/ISADS.1995.398950.

Biba, K. J. (1977). Integrity Considerations for Secure Computer Systems. Online: http://handle.dtic.mil/100.2/ADA039324 [20.06.2008].

Bieberstein, N., S. Bose, L. Walker, and A. Lynch (2005). Impact of Service-oriented Architecture on Enterprise Systems, Organizational Structures, and Individuals. *IBM Systems Journal 44*(4), 691–708. Online: http://www.research.ibm.com/journal/sj/444/bieberstein.pdf [18.11.2008].

Bishop, M. A. (2003). *Computer Security – Art and Science*. Boston: Addison-Wesley.

BITKOM and DIN (2006). Kompass der IT-Sicherheitsstandards - Leitfaden und Nachschlagwerk (Version 2.0). Online: http://www.bitkom.org/files/documents/Kompass_der_IT_28.06.06.pdf [16.11.2006].

Björck, F. (2001). Security Scandinavian Style. Licentiate Thesis, Stockholm University, Department of Computer System Sciences. Online: http://www.dsv.su.se/~bjorck/files/bjorck-thesis.pdf [12.12.2006].

Björck, F. (2004). Institutional Theory: A new Perspective for Research Into IS/IT Security in Organisations. *Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004*, 186–190. DOI: 10.1109/HICSS.2004.1265444.

Blakley, B., E. McDermott, and D. Geer (2001). Information Security is Information Risk Management. In: *NSPW '01: Proceedings of the 2001 Workshop on New Security Paradigms*, New York, pp. 97–104. ACM. DOI: 10.1145/508171.508187.

Böhme, R. (2005). Cyberinsurance Revisited. The Fourth Workshop on the Economics of Information Security (WEIS 2005). Online: http://infosecon.net/workshop/pdf/15.pdf [18.05.2006].

Booker, R. (2006). Re-engineering Enterprise Security. *Computers & Security 25*(1), 13–17. DOI: 10.1016/j.cose.2005.12.005.

Bouchard, M. (2004). The Evolution of Network Security: From DMZ Designs to Devices. META Group Whitepaper. Online: http://www.juniper.net/solutions/literature/white_papers/200084.pdf [20.09.2006].

Bowles, S. (2004). *Microeconomics – Behavior, Institutions, and Evolution*. Princeton, Oxford: Princeton University Press.

Brodt, T. L. and R. M. Verburg (2007). Managing Mobile Work – Insights from European Practice. *New Technology, Work and Employment 22*(1), 52–65.

Brunato, M. and D. Severina (2005). WilmaGate: A new Open Access Gateway for Hotspot Management. In: *Proceedings of the 3rd ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH '05)*, Cologne, pp. 56–64. ACM Press. DOI: 10.1145/1080730.1080740.

Brynjolfsson, E., T. W. Malone, and V. Gurbaxani (1988). Markets, Hierarchies and the Impact of Information Technology. MIT Sloan School Working Paper 2113-88. Online: `http://dspace.mit.edu/retrieve/1848/SWP-2113-27876409.pdf` [18.04.2008].

Brynjolfsson, E., T. W. Malone, V. Gurbaxani, and A. Kambil (1994). Does Information Technology Lead to Smaller Firms? *Management Science 40*(12), 1628–1644.

Brynjolfsson, E. and A. P. McAfee (2007). Beyond Enterprise 2.0. *MIT Sloan Management Review 48*(3), 50–55. Online: `http://sloanreview.mit.edu/smr/issue/2007/spring/03/pdf/48303W.pdf` [05.04.2006].

BSI (2004). *IT Baseline Protection Manual.* Bundesamt für Sicherheit in der Informationstechnik. Online: `http://www.bsi.de/english/gshb/manual/download/pdfversion.zip` [24.09.2006].

BSI (2005a). *BSI Standard 100-1 – Information Security Management Systems (ISMS).* Bundesamt für Sicherheit in der Informationstechnik. Online: `http://www.bsi.de/english/publications/bsi_standards/standard_1001_e.pdf` [05.05.2008].

BSI (2005b). *BSI Standard 100-2 – IT-Grundschutz Methodology.* Bundesamt für Sicherheit in der Informationstechnik. Online: `http://www.bsi.de/english/publications/bsi_standards/standard_1002_e.pdf` [05.05.2008].

BSI (2005c). *BSI Standard 100-3 – Risk Analysis Based on IT-Grundschutz – Version 2.0.* Bundesamt für Sicherheit in der Informationstechnik. Online: `http://www.bsi.de/english/publications/bsi_standards/standard_1003_e.pdf` [05.05.2008].

Camp, L. J. and C. Wolfram (2000). Pricing Security. Proceedings of the CERT Information Survivability Workshop, Boston. Online: `www.cert.org/research/isw/isw2000/papers/54.pdf` [23.02.2007].

Carr, N. G. (2005). The End of Corporate Computing. *MIT Sloan Management Review 46*(3), 67–73. Online: `http://sloanreview.mit.edu/smr/issue/2005/spring/13/` [19.12.2006].

Castells, M. (2000). *Rise of the Network Society: The Information Age: Economy, Society and Culture* (second ed.). Cambridge: Blackwell Publishers.

Cavusoglu, H., H. Cavusoglu, J.-Y. Son, and I. Benbasat (2005). Information Security Control Resources in Organizations: Multidimensionality of the Construct and a Nomological Model. Unpublished Working Paper, University of British Columbia.

Chandran, S. M. and J. Joshi (2005). LoT-RBAC: A Location and Time-Based RBAC Model. In: *Lecture Notes in Computer Science – 6th International Conference on Web Information Systems Engineering 2005*, Volume 3806, pp. 361–375. Berlin, Heidelberg: Springer. DOI: 10.1007/11581062_27.

Cheng, E. K. (2006). Structural Laws and the Puzzle of Regulating Behavior. *Northwestern University Law Review 100*(2), 655–717. Online: `http://www.law.northwestern.edu/journals/lawreview/v100/n2/655/LR100n2Cheng.pdf` [13.08.2008].

Chesbrough, H. W. (2003). The Era of Open Innovation. *MIT Sloan Management Review 44*(3), 35–41. Online: `http://sloanreview.mit.edu/smr/issue/2003/spring/5/` [14.03.2006].

Chia, P. A., S. B. Maynard, and A. B. Ruighaver (2002). Understanding Organizational Security Culture. In: M. G. Hunter and K. K. Dhanda (Eds.), *Information Systems: The Challenges of Theory and Practice*. Las Vegas: Information Institute. Online: `http://www.dis.unimelb.edu.au/staff/sean/research/ChiaCultureChapter.pdf` [18.04.2007].

Ciborra, C. U. (1985). Reframing the Role of Computers in Organizations - The Transaction Costs Approach. In: *Proceedings of the Sixth International Conference on Information Systems*, Indianapolis, pp. 57–69.

Coase, R. H. (1937). The Nature of the Firm. *Economica 4*(16), 386–405.

Coase, R. H. (1960). The Problem of Social Cost. *Journal of Law and Economics 3*(1), 1–44. DOI: 10.1086/466560.

Coase, R. H. (2005). The Institutional Structure of Production. In: C. Ménard and M. M. Shirley (Eds.), *Handbook of New Institutional Economics*, pp. 31–39. Berlin, Heidelberg, New York: Springer.

Cooter, R. (1997). Normative Failure Theory of Law. *Cornell Law Review 82*, 947–979. Online: `http://works.bepress.com/cgi/viewcontent.cgi?article=1020&context=robert_cooter` [27.08.2008].

Cooter, R. (2006). The Intrinsic Value of Obeying Law: Economic Analysis of the Internal Viewpoint. *Fordham Law Review 75*, 1275–1285. Online: `http://law.fordham.edu/publications/articles/500flspub7377.pdf` [27.08.2008].

COSO (1994). Internal Control - Integrated Framework. Committee of Sponsoring Organizations of the Treadway Commission.

Covington, M. J., W. Long, S. Srinivasan, A. K. Dev, M. Ahamad, and G. D. Abowd (2001). Securing Context-Aware Applications Using Environment Roles. In: *Proceedings of the Sixth ACM symposium on Access Control Models and Technologies*, Chantilly, pp. 10–20. ACM. DOI: 10.1145/373256.373258.

Damianides, M. (2005). Sarbanes-Oxley and IT Governance: New Guidance on IT Control and Compliance. *Information Systems Management 22*(1), 77–85. Online: `http://www.infosectoday.com/SOX/Damianides.pdf` [14.11.2006].

Dedrick, J. and K. L. Kraemer (2005). The Impact of IT on Firm and Industry Structure - The Personal Computer Industry. *California Management Review 47*(3), 122–142. Online: `http://www.crito.uci.edu/pubs/2005/impactsOfIT.pdf` [16.04.2008].

Denning, P. J. (2007). Mastering the mess. *Communications of the ACM 50*(4), 21–25. DOI: 10.1145/1232743.1232763.

Denning, P. J. and R. Hayes-Roth (2006). Decision Making in Very Large Networks. *Communications of the ACM 49*(11), 19–23. DOI: 10.1145/1167838.1167852.

Department of Defense (1983/1985). Trusted Computer System Evaluation Criteria. DoD 5200.28-STD. Online: `http://csrc.nist.gov/publications/history/dod85.pdf` [06.12.2006].

Dewett, T. and G. R. Jones (2001). The Role of Information Technology in the Organization: A Review, Model, and Assessment. *Journal of Management 27*(3), 313–346. DOI: 10.1177/014920630102700306.

Dhillon, G. and J. Backhouse (2000). Technical opinion: Information System Security Management in the new Millennium. *Communications of the ACM 43*(7), 125–128. DOI: 10.1145/341852.341877.

Diffie, W. and M. E. Hellman (1976). New Directions in Cryptography. *IEEE Transactions on Information Theory 22*(6), 644–654.

Drucker, P. F. (2006 [1954]). *The Practice of Management* (reissued ed.). New York: HarperCollins. Originally published in 1954.

Drucker, P. F. (2007 [1999]). *Management Challenges for the 21st Century* (revised ed.). Oxford: Butterworth-Heinemann. Originally published in 1999.

Earle, J. S., U. Pagano, and M. Lesi (2002). Information Technology, Organizational Form, and Transition to the Market. Online: `http://ssrn.com/abstract=321202` [15.05.2008]. Upjohn Institute Staff Working Paper No. 02-82.

Eggertsson, T. (1990). *Economic Behavior and Institutions: Principles of Neoinstitutional Economics (Cambridge Surveys of Economic Literature)*. Cambridge: Cambridge University Press.

Ellickson, R. C. (1991). *Order Without Law: How Neighbors Settle Disputes*. Cambridge: Harvard University Press.

Ellickson, R. C. (1999). The Evolution of Social Norms: A Perspective from the Legal Academy. Yale Law School Working Paper # 230. Online: `http://ssrn.com/abstract=191392` [13.08.2008].

European Union (2002). Directive 2002/58/EC of the European Parliament and of the Council Concerning the Processing of Personal Data and the Protection of Privacy in the Electronic Communications Sector (Directive on Privacy and Electronic Communications). Online: `http://europa.eu.int/eur-lex/pri/en/oj/dat/2002/l_201/l_20120020731en00370047.pdf` [03.11.2006].

Eustice, K., L. Kleinrock, S. Markstrum, G. Popek, V. Ramakrishna, and P. Reiher (2003). Securing Nomads: The Case for Quarantine, Examination, and Decontamination. In: *Proceedings of the 2003 Workshop on New Security Paradigms*, Ascona, pp. 123–128. ACM. DOI: 10.1145/986655.986673.

Evans, J. B., W. Wang, and B. J. Ewy (2006). Wireless Networking Security: Open Issues in Trust, Management, Interoperation and Measurement. *International Journal of Security and Networks 1*(1–2), 84–94. DOI: 10.1504/IJSN.2006.010825.

Fehr, E. and S. Gächter (2000). Cooperation and Punishment in Public Goods Experiments. *The American Economic Review 90*(4), 980–994. Online: `http://www.iew.uzh.ch/chairs/fehr/team/fehr/publications/coop_pun.pdf` [29.08.2008].

Feldman, M. S. and J. G. March (1981). Information in Organizations as Signal and Symbol. *Administrative Science Quarterly 26*(2), 171–186. Online: `http://www.jstor.org/stable/2392467` [04.08.2008].

Ferreira, A., R. Cruz-Correia, L. Antunes, P. Farinha, E. Oliveira-Palhares, D. Chadwick, and A. Costa-Pereira (2006). How to Break Access Control in a Controlled Manner. In: *Proceedings of the 19th IEEE International Symposium on Computer-Based Medical Systems*, pp. 847–854. DOI: 10.1109/CBMS.2006.95.

Fluhrer, S., I. Mantin, and A. Shamir (2001). Weaknesses in the Key Scheduling Algorithm of RC4. *Lecture Notes in Computer Science 2259*, 1–24.

Forman, G. H. and J. Zahorjan (1994). The Challenges of Mobile Computing. *IEEE Computer 27*(4), 38–47. DOI: 10.1109/2.274999.

Frey, B. S. and A. Stutzer (Eds.) (2007). *Economics and Psychology: A Promising New Cross-Disciplinary Field*. Cambridge, London: MIT Press.

Friedman, M. (1994 [1953]). The Methodology of Positive Economics. In: D. M. Hausman (Ed.), *The Philosophy of Economics – An Anthology* (second ed.)., pp. 180–213. Cambridge, New York: Cambridge University Press.

Friedman, T. D. (1970). The Authorization Problem in Shared Files. *IBM Systems Journal 9*(4), 258–280. Online: `http://www.research.ibm.com/journal/sj/094/ibmsj0904C.pdf` [18.06.2008].

Furubotn, E. G. and R. Richter (2005). *Institutions & Economic Theory – The Contribution of the New Institutional Economics* (second ed.). Ann Arbor: University of Michigan Press.

Geer, D. E. (2008). *Economics & Strategies of Data Security*. Waltham: Verdasys Press.

Geis, G. S. (2009). The Space Between Markets and Hierarchies. *Virginia Law Review 95*. Forthcoming, preliminary version available online at `http://www.law.virginia.edu/pdf/workshops/0708/geis.pdf` [20.01.2009].

Ghose, A. and U. Rajan (2006). The Economic Impact of Regulatory Information Disclosure on Information Security Investments, Competition, and Social Welfare. The Fifth Workshop on the Economics of Information Security (WEIS 2006). Online: `http://weis2006.econinfosec.org/docs/37.pdf` [22.02.2007].

Glaser, T. (2009). Culture and Information Security – Outsourcing IT Services in China. PhD Thesis, Technical University of Berlin, forthcoming.

Glaser, T. and F. Pallas (2007). Information Security and Knowledge Management: Solutions Through Analogies? *Forschungsberichte der Fakultät IV - Elektrotechnik und Informatik*. Online: `http://ig.cs.tu-berlin.de/ma/fp/ap/2007/GlaserPallas-InformationSecurityAndKnowledgeManagement-2007-09-12.pdf` [02.10.2007].

Glaser, T. and F. Pallas (2008). Aktuelle Sicherheitsparadigmen und ökonomische Konflikte. In: P. Horster (Ed.), *D-A-CH Security 2008 – Bestandsaufnahme, Konzepte, Anwendungen, Perspektiven*, pp. 1–9. Online: `http://ig.cs.tu-berlin.de/ma/fp/ap/2008/` [02.10.2007].

Godber, A. and P. Dasgupta (2002). Secure wireless gateway. In: *Proceedings of the 3rd ACM Workshop on Wireless Security (WiSE '02)*, Atlanta, pp. 41–46. ACM. DOI: 10.1145/570681.570686.

Godber, A. and P. Dasgupta (2003). Countering Rogues in Wireless Networks. In: *Proceedings of the 2003 International Conference on Parallel Processing Workshops*, pp. 425–431. DOI: 10.1109/ICPPW.2003.1240398.

Gordon, L. A., M. P. Loeb, and W. Lucyshyn (2002). An Economics Perspective on the Sharing of Information Related to Security Breaches: Concepts and Empirical Evidence. The First Workshop on the Economics of Information Security (WEIS 2002). Online: `http://www2.sims.berkeley.edu/resources/affiliates/workshops/econsecurity/econws/51.doc` [02.04.2007].

Gould, C., M. Jackson, R. V. Schyndel, and J. O'Donnell (2006). Mapping the Mobile Landscape in Australia. *First Monday 11*(11). Online: `http://firstmonday.org/issues/issue11_11/gould/index.html` [28.11.2006].

Gurbaxani, V. and C. F. Kemerer (1990). An Agency Theory View of the Management of End User Computing. In: *Proceedings of the 11th International Conference on Information Systems*, Copenhagen, pp. 279–290. Online: `http://www.pitt.edu/~ckemerer/CK%20research%`

20papers/AgencyTheoryViewMgtEndUserComputing_GurbaxaniKemerer90.pdf [01.08.2008].

Gurbaxani, V. and S. Whang (1991). The Impact of Information Systems on Organizations and Markets. *Communications of the ACM 34*(1), 59–73. DOI: http://doi.acm.org/10.1145/99977.99990.

Hardin, G. (1968). The Tragedy of the Commons. *Science 162*(3859), 1243–1248. DOI: 10.1126/science.162.3859.1243.

Hennart, J.-F. (1993). Explaining the Swollen Middle: Why Most Transactions are a Mix of "Market" and "Hierarchy". *Organization Science 4*(4), 529–547. Online: http://www.jstor.org/stable/2635079?origin=JSTOR-pdf [08.04.2008].

Henry, P. S. and H. Luo (2002). WiFi: What's Next? *Communications Magazine 40*(12), 66–72. DOI: 10.1109/MCOM.2002.1106162.

Henzel, R. A. (1971). Some Industial Applications of Minicomputers. *IEEE Computer 4*(5), 7–12.

Hinds, P. and S. Kiesler (Eds.) (2002). *Distributed Work*. Cambridge: MIT Press.

HIPAA (1996). Health Insurance Portability and Accountability Act of 1996. One Hundred Fourth Congress of the United States of America, H. R. 3103. Online: http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=104_cong_bills&docid=f:h3103enr.txt.pdf [14.11.2006].

Hodgson, G. M. (2006). What are Institutions? *Journal of Economic Issues 40*(1), 1–25.

Hoffman, L. J. (1977). *Modern Methods for Computer Security and Privacy*. Englewood Cliffs: Prentice-Hall.

Holden, D. (1986). The Role of the Host Computer in Defending Against P.C.s. In: *Proceedings of the Northeast ACM Symposium on Personal Computer Security (PCS '86)*, Waltham, pp. 28–37. ACM. DOI: 10.1145/318772.318780.

Hole, K. J., E. Dyrnes, and P. Thorsheim (2005). Securing Wi-Fi networks. *IEEE Computer 38*, 28–34. DOI: 10.1109/MC.2005.241.

Hong, K., Y. Chi, L. Chao, and J. Tang (2003). An Integrated System Theory of Information Security Management. *Information Management & Computer Security 11*(5), 243–248. DOI: 10.1108/09685220310500153.

Hotspot Directory (2009). The HotSpot Directory – Service Provider Index. Connection Services. Online: http://hotspot-directory.com/site/index.php [26.01.2009].

Housley, R. and W. Arbaugh (2003). Security Problems in 802.11-based Networks. *Communications of the ACM 46*(5), 31–34. DOI: 10.1145/769800.769822.

Huhns, M. and M. P. Singh (2005). Service-oriented Computing: Key Concepts and Principles. *Internet Computing, IEEE 9*(1), 75–81. DOI: 10.1109/MIC.2005.21.

IEEE (2001). 802.1X – Port-Based Network Access Control. IEEE Std 802.1X-2001. Online: `http://standards.ieee.org/getieee802/download/802.1X-2001.pdf` [20.07.2007].

IEEE (2004). 802.11i – IEEE Standard for Information Technology – Telecommunications and Information Exchange Between Systems – Local and Metropolitan Area Networks – Specific Requirements. Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Amendment 6: Medium Access Control (MAC) Security Enhancements. IEEE Std 802.11i-2004. Online: `http://standards.ieee.org/getieee802/download/802.11i-2004.pdf` [20.07.2007].

Isaacson, P., A. Osborne, R. Gammill, L. Tesler, R. Heiser, and J. Warren, J.C. (1978). The Oregon Report – Personal Computing. *IEEE Computer 11*(9), 86–97. DOI: 10.1109/C-M.1978.218349.

ISO / IEC (1998). ISO / IEC 13335-3:1998(E) – Information Technology – Guidelines for the Management of IT Security – Part 3: Techniques for the Management of IT Security. International Organization for Standardization and International Electrotechnical Commission. First edition.

ISO / IEC (2005a). ISO / IEC 17799:2005(E) – Information Technology – Security Techniques – Code of Practice for Information Security Management. International Organization for Standardization and International Electrotechnical Commission. Second edition.

ISO / IEC (2005b). ISO / IEC 27001:2005(E) – Information Technology – Security Techniques – Information Security Management Systems – Requirements. International Organization for Standardization and International Electrotechnical Commission. First edition.

ITGI - The IT Governance Institute (2007). CobiT - Control Objectives for Information and related Technology - Version 4.1. Online: `http://www.isaca.org/cobit` [16.11.2007].

Jaquith, A. (2007). *Security Metrics: Replacing Fear, Uncertainty, and Doubt*. Boston: Addison-Wesley.

Jensen, M. C. and W. H. Meckling (1976). Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics 3*(4), 305–360. Online: `http://ssrn.com/abstract=94043` [14.03.2006].

Kaarst-Brown, M. L. and S. Kelly (2005). IT Governance and Sarbanes-Oxley: The Latest Sales Pitch or Real Challenges for the IT Function? *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05)*. DOI: 10.1109/HICSS.2005.361.

Kallinikos, J. (2007). *The Consequences of Information: Institutional Implications of Technological Change*. Cheltenham, Northampton: Edward Elgar Publishing.

Karyda, M., E. Kiountouzis, and S. Kokolakis (2005). Information Systems Security Policies: A Contextual Perspective. *Computers & Security 24*(3), 246–260. DOI: 10.1016/j.cose.2004.08.011.

Klein, P. G. (2005). The Make-or-Buy Decision: Lessons from Empirical Studies. In: C. Ménard and M. M. Shirley (Eds.), *Handbook of New Institutional Economics*, pp. 435–464. Berlin, Heidelberg, New York: Springer.

Kleinrock, L. (1995). Nomadic Computing – An Opportunity. *SIGCOMM Computer Communication Review 25*(1), 36–40. DOI: 10.1145/205447.205450.

Kleinrock, L. (1997). Nomadic Computing. Online: `http://www.lk.cs.ucla.edu/LK/Bib/PS/paper199.pdf` [10.04.2007].

Kleinrock, L. (2001). Breaking Loose. *Communications of the ACM 44*(9), 41–46. DOI: 10.1145/383694.383705.

Knight, F. H. (2005 [1921]). *Risk, Uncertainty and Profit* (reprinted ed.). Cosimo Publishing. Originally published in 1921.

Knight, W. (2004). Mobile working drives switch to federated access rights. *Infosecurity Today September/October 2004*, 22–25.

Koops, B.-J. (1999). *The Crypto Controversy: A Key Conflict in the Information Society*. The Hague, London, Boston: Kluwer Law International. Online: `http://rechten.uvt.nl/koops/files/page.asp?page_id=21` [11.01.2006].

Kowalski, S. (1994). IT Insecurity: A Multi-disciplinary Inquiry. PhD Thesis, Stockholm University / Royal Institute of Technology, Department of Computer and Systems Sciences, Report No. 94-004. Online: `http://dsv.su.se/en/seclab/pages/pdf-files/94-004.pdf` [12.03.2009].

Kriens, S. (2005). Secure & Assured: Going Beyond the 'Status Quo' Enterprise Network. *Veer: Your Secure and Assured Networking Source* (Spring 2005), 3–4. Online: `http://www.juniper.net/veer/archive/spring_05/pdfs/Veer_Spring_2005.pdf` [20.09.2006].

Kumar, V., R. Telang, and T. Mukhopadhyay (2006). Enterprise Information Security: Who Should Manage it and How? The Fifth Workshop on the Economics of Information Security (WEIS 2006). Online: `http://weis2006.econinfosec.org/docs/21.pdf` [22.06.2006].

Kumar, V., R. Telang, and T. Mukhopadhyay (2007). Optimally Securing Interconnected Information Systems and Assets. The Sixth Workshop on the Economics of Information Security (WEIS 2007). Online: `http://weis2007.econinfosec.org/papers/64.pdf` [29.11.2007].

Kunreuther, H. and G. Heal (2003). Interdependent Security – The Case of Identical Agents. *Journal of Risk and Uncertainty 26*(2), 231–149. DOI: 10.1023/A:1024119208153.

Kuusisto, R., K. Nyberg, and T. Virtanen (2004). Unite Security Culture. May a Unified Security Culture be Plausible? In: A. Jones (Ed.), *Proceedings of the 3rd European Conference on Information Warfare and Security (ECIW 2004)*, London, pp. 191–202. Online: `http://lib.tkk.fi/Diss/2004/isbn9512274639/article7.pdf` [05.09.2008].

Lacey, D. (2005). Inventing the Future – The Vision of the Jericho Forum. *Information Security Technical Report 10*(4), 186–188. DOI: 10.1016/j.istr.2005.10.003.

Laffont, J. and D. Martimort (2001). *The Theory of Incentives - The Principal-Agent Model*. Princeton: Princeton University Press.

Lessig, L. (1998a). The Architecture of Privacy. Online: `http://www.lessig.org/content/articles/works/architecture_priv.pdf` [21.07.2005].

Lessig, L. (1998b). The New Chicago School. *The Journal of Legal Studies 27*(2), 661–691. Online: `http://www.lessig.org/content/articles/works/LessigNewchicschool.pdf` [28.09.2006].

Lessig, L. (1999). *Code and other Laws of Cyberspace*. New York: Basic Books.

Lessig, L. (2006). *Code: Version 2.0*. New York: Basic Books. Online: `http://pdf.codev2.cc/Lessig-Codev2.pdf` [01.08.2008].

Lyytinen, K. and Y. Yoo (2002). Research Commentary: The Next Wave of Nomadic Computing. *Information Systems Research 13*(4), 377–388. DOI: 10.1287/isre.13.4.377.75.

Malone, T. W. (2004). *The Future of Work - How the New Order of Business Will Shape Your Organization, Your Management Style, and Your Life*. Boston: Harvard Business School Press.

Malone, T. W., J. Yates, and R. I. Benjamin (1987). Electronic Markets and Electronic Hierarchies. *Communications of the ACM 30*(6), 484–497. DOI: 10.1145/214762.214766.

Mankiw, N. G. and M. P. Taylor (2006). *Economics*. London: Thomson Learning.

Matsunaga, Y., A. S. Merino, T. Suzuki, and R. H. Katz (2003). Secure Authentication System for Public WLAN Roaming. In: *Proceedings of the 1st ACM International Workshop on Wireless Mobile Applications and Services on WLAN hotspots (WMASH '03)*, San Diego, pp. 113–121. ACM. DOI: 10.1145/941326.941343.

Mavridis, I. and G. Pangalos (1997). Security Issues in a Mobile Computing Paradigm. In: *Proceedings of Communications and Multimedia Security (CMS'97)*, Volume 3, pp. 60–76.

Ménard, C. (2005). A New Institutional Approach to Organization. In: C. Ménard and M. M. Shirley (Eds.), *Handbook of New Institutional Economics*, pp. 281–318. Berlin, Heidelberg, New York: Springer.

Ménard, C. and M. M. Shirley (Eds.) (2005). *Handbook of New Institutional Economics*. Berlin, Heidelberg, New York: Springer.

Mercuri, R. T. (2003). Analyzing Security Costs. *Communications of the ACM 46*(6), 15–18. DOI: 10.1145/777313.777327.

Milgrom, P. and J. Roberts (1992). *Economics, Organization & Management*. Upper Saddle River: Prentice Hall.

Murray, W. H. (1984). Security Considerations for Personal Computers. *IBM Systems Journal 23*(3), 297–304. Online: `http://www.research.ibm.com/journal/sj/233/ibmsj2303I.pdf` [10.07.2008].

Murray, W. H. (1986). Good security practice for personal computers. In: *Proceedings of the Northeast ACM Symposium on Personal Computer Security (PCS '86)*, Waltham, pp. 1–12. ACM. DOI: 10.1145/318772.318775.

Nisan, N. (2007). Introduction to Mechanism Design (for Computer Scientists). In: N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani (Eds.), *Algorithmic Game Theory*, pp. 209–242. Cambridge, New York: Cambridge University Press.

NIST (2002). Risk Management Guide for Information Technology Systems. Online: `http://www-08.nist.gov/publications/nistpubs/800-30/sp800-30.pdf` [06.05.2008]. National Institute of Standards and Technology, Special Publication 800-30.

Nooteboom, B. (1992). Information Technology, Transaction Costs and the Decision to 'Make or Buy'. *Technology Analysis & Strategic Management 4*, 339–350. DOI: 10.1080/09537329208524105.

North, D. C. (1990). *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press.

OECD (1992). Guidelines for the Security of Information Systems. Organisation for Economic Co-operation and Development. Online: `http://www.oecd.org/document/19/0,2340,en_2649_34255_1815059_1_1_1_1,00.html` [13.03.2009].

OECD (2002). Guidelines for the Security of Information Systems and Networks. Towards a Culture of Security. Organisation for Economic Co-operation and Development. Online: `http://www.oecd.org/dataoecd/16/22/15582260.pdf` [13.05.2006].

Ostrom, E. (2005). *Understanding Institutional Diversity*. Princeton: Princeton University Press.

Padayachee, K. and J. H. P. Eloff (2007). Enhancing Optimistic Access Controls with Usage Control. In: *Lecture Notes in Computer Science – Trust, Privacy and Security in Digital Business 2007*, Volume 4657, pp. 75–82. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-540-74409-2_10.

Palmer, G. (2005). De-Perimeterisation: Benefits and Limitations. *Information Security Technical Report 10*(4), 189–203. DOI: 10.1016/j.istr.2005.09.001.

Panagacos, J. T., A. R. Farber, S. M. Dreyer, and D. H. McVeigh (1985). Managing Computing Resources (Panel Session): The Personal Computer Revolution. In: *Proceedings of the 1985 ACM Annual Conference on the Range of Computing: Mid-80's Perspective*, Denver, pp. 350–351. ACM. DOI: 10.1145/320435.320538.

Papazoglou, M. P. and D. Georgakopoulos (2003). Service-Oriented Computing – Introduction. *Communications of the ACM 46*(10), 25–28. DOI: 10.1145/944217.944233.

PCAOB - Public Company Accounting Oversight Board (2004). Board Considers Adopting Standard for Audits of Internal Control Over Financial Reporting – Briefing Paper. Online: `http://www.pcaob.com/Rules/Docket_008/2004-03-09_Briefing_Paper.pdf` [23.11.2006].

PCAOB - Public Company Accounting Oversight Board (2007). Auditing Standard No. 5 – An Audit of Internal Control Over Financial Reporting That is Integrated With an Audit of Financial Statements and Related Independence Rule and Conforming Amendments. Online: `http://www.pcaobus.org/Rules/Docket_021/2007-06-12_Release_No_2007-005A.pdf` [16.06.2008].

Peltier, T. R. (2005). *Information Security Risk Analysis* (second ed.). Boca Raton: Auerbach Publications / CRC Press.

Perry, M., K. O'Hara, A. Sellen, B. Brown, and R. Harper (2001). Dealing with Mobility: Understanding Access Anytime, Anywhere. *ACM Transactions on Computer-Human Interaction 8*(4), 323–347. DOI: 10.1145/504704.504707.

Picot, A., H. Dietl, and E. Franck (2005). *Organisation. Eine ökonomische Perspektive* (fourth ed.). Schäffer-Poeschel.

Pinder, P. (2006). Preparing Information Security for Legal and Regulatory Compliance (Sarbanes-Oxley and Basel II). *Information Security Technical Report 11*(1), 32–38. DOI: 10.1016/j.istr.2005.12.003.

Pinsonneault, A. and K. L. Kraemer (2002). Exploring the Role of Information Technology in Organizational Downsizing: A Tale of Two American Cities. *Organization Science 13*(2), 191–208. DOI: 10.1287/orsc.13.2.191.537.

Png, I., C. Q. Tang, and Q.-H. Wang (2006). Hackers, Users, Information Security. The Fifth Workshop on the Economics of Information Security (WEIS 2006). Online: `http://weis2006.econinfosec.org/docs/54.pdf` [22.06.2006].

Posner, R. A. (1978). The Right of Privacy. *Georgia Law Review 12*(3), 393–422. Online: `http://digitalcommons.law.uga.edu/lectures_pre_arch_lectures_sibley/22/` [09.02.2009].

Posner, R. A. (1997). Social Norms and the Law: An Economic Approach. *The American Economic Review 87*(2), 365–369.

Posner, R. A. and E. B. Rasmusen (1999). Creating and Enforcing Norms, with Special Reference to Sanctions. *International Review of Law and Economics 19*, 369–382. DOI: 10.1016/S0144-8188(99)00013-7.

Potter, B. (2006). Wireless Hotspots: Petri Dish of Wireless Security. *Communications of the ACM 49*(6), 50–56. DOI: 10.1145/1132469.1132501.

Povey, D. (2000). Optimistic Security: A New Access Control Paradigm. In: *Proceedings of the 1999 Workshop on New Security Paradigms*, Caledon Hills, pp. 40–45. ACM. DOI: 10.1145/335169.335188.

Powell, W. W. (1987). Hybrid Organizational Arrangements: New Form or Transitional Development. *California Management Review 30*(1), 67–87.

Pretschner, A., F. Massacci, and M. Hilty (2007). Usage Control in Service-Oriented Architectures. In: *Lecture Notes in Computer Science – Trust, Privacy and Security in Digital Business 2007*, Volume 4657, pp. 83–93. Berlin, Heidelberg: Springer. DOI: 10.1007/978-3-540-74409-2_11.

Quillard, J. A., J. F. Rockart, E. Wilde, M. Vernon, and G. Mock (1983). A Study of Corporate Use of Personal Computers. *Sloan Working Papers*. Online: `http://dspace.mit.edu/handle/1721.1/2066` [01.08.2008].

Reidenberg, J. R. (1998). Lex Informatica: The Formulation of Information Policy Rules Through Technology. *Texas Law Review 76*(3), 553–584. Online: `http://reidenberg.home.sprynet.com/lex_informatica.pdf` [27.09.2006].

Ricardo, D. (2002 [1817]). *The Principles of Political Economy and Taxation* (third ed.). Janus Publishing. Originally published in 1817.

Rissanen, E., B. S. Firozabadi, and M. Sergot (2006). Towards A Mechanism for Discretionary Overriding of Access Control. *Lecture Notes in Computer Science 3957*, 312–319. DOI: 10.1007/11861386.

Rockart, J. F. and L. S. Flannery (1983). The Management of End User Computing. *Communications of the ACM 26*(10), 776–784. DOI: 10.1145/358413.358429.

Rowe, B. R. and M. P. Gallaher (2006). Private sector Cyber Security Investment Strategies: An Empirical Analysis. *The Fifth Workshop on the Economics of Information Security (WEIS 2006)*. Online: `http://weis2006.econinfosec.org/docs/18.pdf` [03.07.2006].

Russell, D. and G. T. Gangemi (1991). *Computer Security Basics*. Sebastopol: O'Reilly.

Saltzer, J. H. (1974). Protection and the Control of Information Sharing in Multics. *Communications of the ACM 17*(7), 388–402. DOI: 10.1145/361011.361067.

Sandhu, R. S., E. J. Coyne, H. L. Feinstein, and C. E. Youman (1996). Role-Based Access Control Models. *IEEE Computer 29*(2), 38–47. DOI: 10.1109/2.485845.

SANS (2006). The SANS Security Policy Project - Email Use Policy - Template. Online: `http://www.sans.org/resources/policies/Email_Policy.pdf` [16.02.2009].

Sappington, D. E. M. (1991). Incentives in Principal-Agent Relationships. *The Journal of Economic Perspectives 5*(2), 45–66. Online: `http://www.jstor.org/stable/1942685` [11.08.2008].

Schlienger, T. and S. Teufel (2002). Information Security Culture: The Socio-Cultural Dimension in Information Security Management. In: *Proceedings of the IFIP TC11 17th International Conference on Information Security (SEC '02)*, Deventer, pp. 191–202. Kluwer.

Schneier, B. (2000). *Secrets & Lies – Digital Security in a Networked World*. Indianapolis: John Wiley & Sons.

Schneier, B. (2002). Computer Security: It's the Economics, Stupid. The First Workshop on the Economics of Information Security (WEIS 2002). Online: `http://www2.sims.berkeley.edu/resources/affiliates/workshops/econsecurity/econws/18.doc` [23.02.2009].

Schneier, B. (2006). *Beyond Fear*. Berlin, Heidelberg, New York: Springer.

Schneier, B. (2007a). How Security Companies Sucker Us With Lemons. Wired Online. Online: `http://www.wired.com/politics/security/commentary/securitymatters/2007/04/securitymatters_0419?currentPage=all` [20.04.2007].

Schneier, B. (2007b). Information Security and Externalities. *ENISA Quarterly 2*(4), 3–4. Online: `http://www.enisa.europa.eu/doc/pdf/publications/enisa_quarterly_01_07.pdf` [18.01.2007].

Schneier, B. (2007c). Why Smart Cops Do Dumb Things. Wired Online. Online: `http://www.wired.com/politics/security/commentary/securitymatters/2007/02/72774?currentPage=all` [04.08.2008].

Schneier, B. (2008). Security ROI: Fact or Fiction? CSO Online. Online: `http://www.csoonline.com/article/446866/Security_ROI_Fact_or_Fiction_` [04.10.2008].

Scott, W. R. (2001). *Institutions and Organizations* (second ed.). Foundations for Organizational Science. Sage Publications.

Shamir, A. (2002). Cryptography: State of the Science – Turing Lecture 2002. Online: `http://awards.acm.org/images/awards/140/vstream/2002/S/s-pp/shamir_1files.html` [28.04.2008].

Shapiro, C. and H. Varian (1999). *Information Rules.* Boston: Harvard Business School Press.

Sheldon, T. (2001). *Encyclopaedia of Networking & Communications.* New York, Chicago, San Francisco: McGraw-Hill.

Shepherd, M. (1977). Distributed Computing Power: A Key to Productivity. *IEEE Computer 10*(11), 66–74. DOI: 10.1109/C-M.1977.217568.

Shin, N. (1997). The Impact of Information Technology on Coordination Costs: Implications for Firm Productivity. In: *Proceedings of the Eighteenth International Conference on Information Systems (ICIS '97)*, Atlanta, pp. 133–146. Association for Information Systems.

Shostack, A. and A. Stewart (2008). *The New School of Information Security.* Upper Saddle River: Addison-Wesley.

Singh, S. (1999). *The Code Book – The Secret History of Codes and Code-Breaking.* London: Fourth Estate.

Siponen, M. T. (2000). A Conceptual Foundation for Organizational Information Security Awareness. *Information Management & Computer Security 8*(1), 31–41. DOI: 10.1108/09685220010371394.

Siponen, M. T. (2006a). Information Security Standards Focus on the Existence of Process, not its Content. *Communications of the ACM 49*(8), 97–100. DOI: 10.1145/1145287.1145316.

Siponen, M. T. (2006b). Secure-System Design Methods: Evolution and Future Directions. *IT Professional 8*(3), 40–44. DOI: 10.1109/MITP.2006.73.

Smith, A. (1982 [1776]). *An Inquiry into the Nature and Causes of the Wealth of Nations* (reprinted ed.). Penguin Classics. Edited by Andrew Skinner, originally published in 1776.

SOX (2002). Sarbanes-Oxley Act of 2002 – Public Company Accounting Reform and Investor Protection Act of 2002. One Hundred Seventh Congress of the United States of America, H. R. 3763. Online: `http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_cong_bills&docid=f:h3763enr.txt.pdf` [07.11.2006].

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics 87*(3), 355–374. DOI: 10.2307/1882010.

Steinauer, D. D. (1985). Security of Personal Computer Systems: A Management Guide. Technical Report NBS-SP-500-120, National Bureau of Standards (DOC), Institute for Computer Sciences and Technology, Washington, DC. Online: http://eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED274325 [31.07.2008].

Stiglitz, J. E. (2008). Principal and Agent (ii). In: S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics Online* (second ed.). Palgrave Macmillan. DOI: 10.1057/9780230226203.1342.

Straub, D. W. (1990). Effective IS Security: An Empirical Study. *Information Systems Research 1* (3), 255–276.

Stubblefield, A., J. Ioannidis, and A. D. Rubin (2002). Using the Fluhrer, Mantin, and Shamir Attack to Break WEP. Network and Distributed System Security Symposium 2002. Online: http://www.isoc.org/isoc/conferences/ndss/02/proceedings/papers/stubbl.pdf [13.02.2007].

Surden, H. (2007). Structural Rights in Privacy. *SMU Law Review 60*, 101–145. Online: http://ssrn.com/abstract=1004675 [13.08.2008].

Tews, E., R.-P. Weinmann, and A. Pyshkin (2007). Breaking 104 Bit WEP in Less Than 60 Seconds. Cryptology ePrint Archive, Report 2007/120. Online: http://eprint.iacr.org/cgi-bin/cite.pl?entry=2007/120 [16.04.2007].

Thomas, R. C. and P. D. Amon (2007). Incentive-based Cyber Trust – A Call to Action. Online: http://meritology.com/resources/Incentive-based%20Cyber%20Trust%20Initiative%20v3.5.pdf [03.06.2008].

Varian, H. (2004). System Reliability and Free Riding. Online: http://www.sims.berkeley.edu/~hal/Papers/2004/reliability.pdf [14.03.2006].

Voigt, S. (2002). *Institutionenökonomik.* München: Utb / W. Fink.

von Hayek, F. A. (1945). The Use of Knowledge in Society. *The American Economic Review 35* (4), 519–530.

von Solms, B. (2000). Information Security - The Third Wave? *Computers & Security 19*, 615–620. DOI: 10.1016/S0167-4048(00)07021-8.

VPN Consortium (2006). VPN Technologies: Definitions and Requirements. Online: http://www.vpnc.org/vpn-technologies.html [12.07.2007].

Ware, W. H. (1970). Security Controls for Computer Systems. Published by the RAND Corporation – Defense Science Board – Task Force on Computer Security. Online: http://csrc.nist.gov/publications/history/ware70.pdf [30.08.2006].

Whitman, M. E. and H. J. Mattord (2003). *Principles of Information Security.* Beijing: Thomson Learning. Reprint by Tsinghua University Press.

Wi-Fi Alliance (2003a). Best Current Practices for Wireless Internet Service Provider (WISP) Roaming. Document only available via secondary resources.

Wi-Fi Alliance (2003b). Enterprise Solutions for Wireless LAN Security. Online: `http://www.wi-fi.org/files/wp_3_Securing%20Wi-Fi%20In%20The%20Enterprise_2-6-03.pdf` [16.02.2007].

Wi-Fi Alliance (2004a). Enabling the Future of Wi-Fi Public Access. Online: `http://www.wi-fi.org/files/wp_2_Future%20of%20Wi-Fi%20Public%20Access_1-2-04.pdf` [04.07.2007].

Wi-Fi Alliance (2004b). WPA Deployment Guidelines for Public Access Wi-Fi Networks. Online: `http://www.wi-fi.org/files/wp_6_WPA%20Deployment%20for%20Public%20Access_10-28-04.pdf` [16.02.2007].

Wiant, T. L. (2005). Information Security Policy's Impact on Reporting Security Incidents. *Computers & Security 24*, 448–459. DOI: 10.1016/j.cose.2005.03.008.

Williams, J. (2002). Providing for Wireless LAN Security, Part 2. *IT Professional 4* (6), 44–48. DOI: 10.1109/MITP.2002.1114847.

Williamson, O. E. (1975). *Markets and Hierarchies – Analysis and Antitrust Implications*. New York: Free Press.

Williamson, O. E. (1985). *The Economic Institutions of Capitalism*. New York: Free Press.

Williamson, O. E. (1991). Comparative Economic Organization: The Analysis of Discrete Structural Alternatives. *Administrative Science Quarterly 36* (2), 269–296.

Williamson, O. E. (2005). Transaction Cost Economics. In: C. Ménard and M. M. Shirley (Eds.), *Handbook of New Institutional Economics*, pp. 41–65. Berlin, Heidelberg, New York: Springer.

Yates, J., W. J. Orlikowski, and S. L. Woerner (2003). Virtual Organizing: Using Threads to Coordinate Distributed Work. In: *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, p. 271.2. IEEE. DOI: 10.1109/HICSS.2003.1174796.

Yngström, L. (1996). A Systemic-Holistic Approach to Academic Programmes in IT Security. PhD Thesis, Stockholm University / Royal Institute of Technology, Department of Computer and Systems Sciences, Report No. 96-021. Online: `http://dsv.su.se/en/seclab/pages/pdf-files/96-021.pdf` [12.03.2009].

Zhang, G. and M. Parashar (2003). Dynamic context-aware access control for grid applications. In: *Proceedings of the Fourth International Workshop on Grid Computing*, pp. 101–108. IEEE. A slightly different version of the paper is also available at `http://www.caip.rutgers.edu/TASSL/Papers/automate-sesame-cnds-04.pdf` [16.01.2009].

Zhao, X. and M. E. Johnson (2008). Information Governance: Flexibility and Control through Escalation and Incentives. The Seventh Workshop on the Economics of Information Security (WEIS 2008). Online: `http://weis2008.econinfosec.org/papers/Zhao.pdf` [11.02.2009].

Zimmermann, H. (1980). OSI Reference Model–The ISO Model of Architecture for Open Systems Interconnection. *IEEE Transactions on Communications 28*(4), 425–432.