Security: always too much and never enough Anthropology of a non-starter market

Dominique BOULLIER*, Pascal JOLLIVET**, Frédéric AUDREN***

Abstract

The security market, based on public Key Infrastructures (PKI) did not succeed because security remains a paradoxical market. We observed security practices and reciprocal expectations, in this study the ones generated by the design of PKI devices. Using the framework of Actor Network Theory, we describe all the mediations required for sustaining a digital security chain… often based on very material stuff. A whole vision of the world should be designed, an ontology, doomed to failure if it formats practices and users by constraint. This vision should retain a variable-geometry, while calling on guarantors that transcend it, and not merely on commercial certification authorities. Will security architecture design be able to integrate the users' demand for "adequate security", which renders security policies bearable as long as users are not aware of them?

Key words: PKI, Actor Network Theory, Digital networks security, Mediations, Socio-technical architecture, Market composition, Acceptability, Users study.

TITRE FRANÇAIS

Résumé

Le marché de la sécurité sous forme des Public Key Infrastructures (Infrastructure de gestion de clés) n'est pas parvenu à décoller car la sécurité reste un marché paradoxal. Nous avons observé les pratiques de sécurité et les attentes réciproques créées par la conception de ces systèmes, plus spécifiquement ceux à base de PKI pour cette étude, dans les termes de la théorie de l'acteur-réseau, en reconstituant toutes les médiations nécessaires à l'existence d'une chaîne de sécurité informatique… souvent bien matérielle. C'est une vision sécuritaire du monde qui doit être produite, une ontologie, qui échoue quand elle veut trop formater les pratiques et les utilisateurs : elle doit rester « à géométrie variable » tout en mobilisant des garants qui la dépassent et non les seules autorités de certification marchandes. La conception d'architectures de sécurité peut elle admettre cette « sécurité suffisante » qui rend sup-

* Université Rennes 2, LAS EA 2241 – ZAC Atalante Champeaux ; 3, allée Adolphe Bobierre, 35000 Rennes, France.
** Université de Technologie de Compiègne, Costech EA2223 – Rue Roger Couttolenc, 60206 Compiègne cedex, France.
*** CNRS, *Adresse ?*

portable les politiques de sécurité dès lors qu'elles disparaissent de la conscience des utilisateurs ?

Mots clés : PKI, Théorie de l'acteur réseau, Sécurité informatique, Médiations, Architecture socio-technique, Composition de marché, Acceptabilité, Étude d'usages.

Contents

I. *Introduction*
II. *Designing a security chain*
III. *Looking for the lost guarantor: "the body natural" or the "body politick"*
IV. *Security: the never-ending quest*
V. *Bringing a "non-starter" market to life*
VI. *Conclusion*
VII. *Creating a market for a public good?*
References (29 ref.)

I. INTRODUCTION

Ordinary users ask for greater security for their document transfers, for their personal data, for electronic transactions, etc. but they want it to be done unobtrusively, seamlessly and without having to change their own ways. But however efficient security software and system architectures prove to be, they cannot operate without some degree of daily implication on the part of the users to hold the whole security chain together. So why is it that a "need" identified in any number of questionnaires and marketing surveys does not translate into effective action? What is the point of "needs analysis" today, when we know that the technology exists and that the main issue is about understanding why so few people are interested in using the products available? This is precisely why we never talk about "needs": there is no "need for security" and the very notion is meaningless, even when it is discussed in more concrete terms, for instance in the guise of electronic signatures. Network security systems such as PKIs (Public Key Infrastructures) are of course both readily available and technically reliable, but people are still reluctant to use them. Despite all the media interest, demand for such systems remains desperately low.

Social scientists can help examine the whole concept of demand for security in order to understand why a seemingly uncontroversial product or concept like IT security has failed miserably as a commercial and practical proposition. We therefore have to rephrase the security question and adopt a radically different methodological approach in order to discover what is meaningful for users on a daily basis and state the qualities that users might be looking for in security chain components, that could form the basis for a stable and attractive security solution.

The study we conducted[1] (Audren, Boullier, Jollivet, 2002), in such a context of a "sinking" innovation, is necessarily a complex one and is inevitably incomplete, but it should help us understand what is at stake in the digital data exchange security issue. Understanding

1. These observations were carried out under a French RNRT "I-Care" research project ("Trust infrastructure in internet and mobile phone networks"), undertaken between September 2001 and September 2003. See "methodology".

means taking it as a given that the behaviour of the operators under observation is rational, in the sense that everyone can come up with good reasons why they do not comply with the very security rules that they otherwise wholeheartedly support. The design process itself and the type of people involved needs to be revised to ensure a holistic approach to system security where every component in the system, whether it is a human operator, a software application, a computer terminal or a regulation, is assigned a vital role.

This is where Actor Network Theory (Latour, 2005) can be a useful framework, preventing us from assigning, a priori, major or minor roles, social or technical ones, to entities that would only appear in the course of the development as crucial ones, acting as "mediations" and not as tools anymore. Composing a market (Callon, 1998) for security is not a trivial task since it does not merely *require* trust, something familiar to economists, but il *sells* trust, which is a volatile good, more akin to the reputation that nowadays fuels the financial economy. All stakeholders are looking for the best way to display some material cues in order to enhance the reality of this good. We shall demonstrate how this market can be composed with tiny bits of paper that can disrupt the security chain on one hand, and on large scale ontologies on the other hand, asserting who, in this uncertain world, is the final trustable entity that allows us to stop doubting our own identities. The security market must control both approaches and many other forms of mediation if it is to exist. The paradox we would like to emphasize is that it requires a full reversal of method for system designers who have to get rid of their fully deployed and mandatory charts and procedures in any kind of organization. It becomes a much more opportunistic job where developers follow acceptability level and security routines from the user point of view and wait for circumstances that would allow them to improve their systems at the very moment when the crisis and security failures force users to revise (Livet, 1993) their previous conventions (Orléan, 1994; Eymard-Duvernay and al., 2004).

This survey follows the approach developed by Actor Network Theory sociology (ANT)[2] (Callon, 2001) and gives prominence to the specific physionomy of technical objects. While it is all too often supposed that sciences express a form of truth which is not subject to social contingencies, technical entities are, conversely, deemed composite and heterogeneous by nature. "A technical object is the formalisation and measurement of a set of relations between completely disparate elements" (Akrich, 1987, p. 160). The technical object plays an active part in building up networks linking together actants which are dissimilar by nature and size, which need to be precisely described. There is therefore no fundamental divide between society and technology. More precisely, both are defined concurrently in one seamless continuum (Latour, 1990). In this perspective, understanding a security system implies a knowledge of the specific local social fabric, an understanding of what makes it fertile, as well as a capacity to grasp the complexity of the mediation processes on which its formal stability is based (Latour, 2006). The very definition of security is problematic, uncertain and controversial when it comes to describing its possible architectures in the minutest detail. Only by tracing these minute inflexions, subtle doubts and ambiguities can one hope to understand how a security policy finally emerges. The policy emerges as a result, not of strict compliance with any formal definitions which may have been given, but by a series of gradual adjustments leading eventually to a stable system. Like "society", "security" is therefore neither result nor cause but something that emerges above and beyond its constituent conditions and elements.

2. Traduction de : Actor-Network theory.

II. DESIGNING A SECURITY CHAIN

The variety of means implemented to ensure security is often only mentioned as an afterthought, which is why systems such as PKI are often seen as belonging to a remote and artificially created universe. Yet the quality of the relations between the PKI and all the different mediators in the system determines the quality of the security chain itself. We must list what designing this chain should include:

- physical and digital items;
- spaces, networks, contents and people;
- identified enemies and suspects;
- boundaries between in-group and out-group;
- target populations

This list demonstrates as such what is a "security policy", i.e. a whole new way of understanding one's world through the lens of security requirements, that consider every entity as a potential risk.

II.1. The physical and the digital: a question of detail

It would be tempting to try to construct models of corporate security systems or information system security in general, or to focus only on what a developer would be interested in, i.e. network security. But the developer is in fact influenced by the very conception of the information system as a "computer" system, whereas no information system can disregard the "physical" context in which it operates, including the material elements of the system themselves (the actual machines and interfaces). Conversion from one world to the other (i.e. from the intangible to the tangible) is not something that simply takes place at one end of the chain and is reversed at the other: any information system is an intricate construction involving software and hardware but also technical and organisational choices. At any given moment, every operator carries out multiple two-way conversions from the physical to the digital, from hardware to software. Passwords are typical of the kind of supposedly secure component which is in fact a weak link in the security chain because it is often jotted down on a paper sticker or carelessly passed on to colleagues. Behaviour of this kind is always meaningful and can be justified, but it is not generally taken into account in the security chain, or is simply considered deviant whereas it is in fact quite common and "normal". Access keys, whether physical or virtual in nature (i.e. badges, codes, etc.) raise the same kind problem. Allowing access to the premises where the machines and/or the data that needs protecting are located means that modelisation cannot be reduced to the computer system per se. This necessarily involves different types of conversions and exchanges which are going to be as many weak links in the security chain. You can of course try to identify visible evidence of security-awareness, such as safes or briefcases that people always carry around with them, but it is virtually impossible to take all these links in the chain into account. The question of both physical and digital access controls is a central one which sometimes seems to encompass the whole security policy issue. The computerized model of security lacks the

adaptative ability necessary for considering these objects (badges, codes and so on) as mediations and not only as tools, i.e.able to become key factors for the conversion of the whole security chain in an effective system.

These objects seem typical of what mediations can do: they really contribute to security and to change the world they are acting in while being dealt as simple tools, i. e. "slaves" or "non autonomous entities" (Latour, 1994). But their mediation status can only appear clearly when the situation goes wrong, when it appears that these "natural objects" do not merely "function" anymore. The naturalisation of these tiny objects, as well as of the procedures or large institutions, is required in order to create a safe environment, i.e. felt as safe and naturally safe, unquestionable, implicit. Things work at their best when they let us feel as if they do not exist. But it is this very feeling that puts security at risk: neglecting their role in the practical chain of security, or in the design process of PKI as well, makes room for security failures. This has been quite well documented in man machine interactions studies, describing the contradiction between acting as much naturally as possible for efficiency and keeping one aware of change cues for safety control at the same time (Rasmussen, 1986; Reason, 1990).

II.2. Of space, networks, contents, and people

Security must be implemented in all the entities in the organisation, whatever their various manifestations (physical or logic-based) within the information system. Physical space comprises all the walls, doors, windows, etc. which can be secured by alarms, keys and so on. Networks consist of wires, waves, algorithms, servers, data flows, etc. and are secured by passwords, various kinds of scrambling algorithms, and keys to access the machines themselves. Contents are made up of languages, images, digits, titles, indexing parameters and digitized documents and are secured by prints, encryption systems, signatures and logins or passwords. People translate into flesh and blood, sensory modes, institutionally recognised identities and intentions and can be secured by fingerprints, identification documents, physical barriers, surveillance and behaviour tracking, etc. These persons can be categorized according to function or project, which implies a totally different approach to security issues. When their definition is function-based, their status can be considered as stable, yet the level of security applied to them may vary according to the projects in which they are involved. When their definition is project-based, their status is by definition temporary and ad hoc, and will need to be regulary updated.

As we said previously, all it takes is for one of these entities or one of their manifestations to be overlooked, or their intentions or behavioural patterns to be wrongly predicated, and the whole security chain can be jeopardized. The first priority in attempting to deal with security issues should therefore be to define an extensive, distributed and multifarious ontology, if one follows the usual reliability criteria and the AI experts' natural tendency to start by modelizing every possible situation. Anyone mad enough to take on this task, however, would spend a whole lifetime at it and would be incapable of dealing with the changes and uncertainties that characterize organisations in general.

II.3. Inside and out: the enemy within and the enemy beyond

It goes without saying that it is impossible to reify the borderline between inside and out when it comes to defining technical objects. The borderline is more the result of interaction between the technical object itself, its different usages and the actors involved, than a clearly defined beginning. Nevertheless, a security policy should clearly focus on the adversaries that it sets out to identify and put out of action. This may seem simple enough when setting up firewalls to deal with the kind of external attacks that all organisations and most individual users have had to deal with. Everyone is aware that evildoers are lurking on the network. The only visible trace is often a computer virus, but this can be sufficiently painful to make people put up with counter-measures.

Nowadays, the threat is more likely to be related to information property or privacy, and security has to be guaranteed in these areas as well. However, this touches on home-grown enemies, who might be both neighbours and competitors; in other words, anyone who might have an interest in using the coveted information. In this case, everyone becomes a potential suspect and the security system has to turn in to an internal surveillance system monitoring the behaviour of anyone within the institution. The systems engineer or administrator himself can then become the prime suspect in the eyes of the people under surveillance, because he has access to even the most intimate data for technical reasons. Of course, the general assumption that his professional ethics will help him steer clear of any wrongdoing prevents everyone from becoming permanently paranoid, but "you never know", what he might be capable of with all that freedom of access… After all, the basic question of who controls the controller is a vital issue in any human system, be it security-based or otherwise!

II.4. Target populations: when computer scientists try to generate clones

In defining the different types of people targeted by security systems, a kind of stereotyped formatting takes place. The formatting process is predicated on the idea that it is the users who must adapt to the technology and not the opposite.

This "user formatting process" is conducted first of all via the user training and assistance support materials such as paper or online documentation, and various other aids designed to tie a user to a given application. Contrary to a widely shared view, users are in fact ready to go to remarkable lengths to adapt their ways of doing things, providing they are given the opportunity to make the changeover once and for all and be given clear, decisive and reliable instructions (Boullier, 2001).

The second type of user formatting takes place via the definition of sub-groups of users by the security system being implemented. Some users will thus be included in the "inner circle" while others will be excluded, both from certain types of services and from certain phases of the deployment cycle. As regards deployment, accepted practice is for the initial target population to include the *IT specialists*: this is likely to result in norm-building and standardisation which will not be conducive to the adoption and use of the system by other *ordinary* users. Simply studying these sub-groups is not very useful as this would result in reducing them to a single identity with a set of properties, whereas the transversal properties common to all the groups will need to be taken into account if one wants to define different

profiles. Identifying the *attributes* of all the different user populations would on the contrary allow security measures to be fine-tuned by implementing specific, systematic training schemes and through awareness-raising material.

It would be useful to undertake such a systematic description of the properties and attributes of the user populations during the actual development phase and well before implementation, as implicit options tend to automatically translate into system architecture choices which will be very difficult to modify at a later stage. Defining target population attributes means that pre-determined internal borderlines have to be redrawn, while hierarchies and implicit roles and organisational and network governance principles have to be redefined. These need to be solidly defined before they can be deployed in specific instances.

III. LOOKING FOR THE LOST GUARANTOR: THE "BODY NATURAL" OR THE "BODY POLITICK" [3]

The world designed by a security policy seems already quite well populated as we demonstrated with this list. Nevertheless, the security quest requires a more institutionalized status, i.e. a way of keeping together the personal and the global, in order to produce a fully accountable ontology. Strangely enough, it will rely on the same cues of guarantee that the first designers of identities in modern states had to trust. This quest for a guarantor is the corner stone of any security policy (which has to define one or several "authorities", i.e. a guarantor who is both "human" while standing above common humanity). This quest cannot be overlooked because it lies at the very heart of the institution and of the beliefs that hold it together (Legendre, 1983). This is why any security policy has to provide reassurances, to allay suspicions about technical mediation and the individuals in charge. Information system security premises can never be self-justifying: they should always refer to something greater than the system itself and appear to be beyond the reach of any possible digital manipulation. The two references from outside the information system that were most frequently mentioned by respondents who recognize the need for an external guarantor were:

- the human body, as exemplified in the current "holy grail" of biometrics and visual authentification, and
- the State, as the supposed source of all authority.

Here we have the two opposite poles on which present-day beliefs are predicated: the individual, supposedly identifiable with his or her biological properties, or so science would have us believe; and the State, which is supposed to transcend both its originators and political parties and social groups, according to the political tenets of the modern nation-state. Both these answers illustrate the need to free the issue of any partisan controversy by returning to basic and incontrovertible truths. The actors no longer know where they stand and are torn between the model of the sovereign State and that of a society dominated by biopolitical control (Deleuze, 1990). Truths and identities that digitisation has fundamentally brought into question by enabling all manner of manipulation.

3. A distinction introduced by Samuel Johnson during the XVIIIth century refering to John Locke theory.

Biometrics on the one hand encourages an ever-increasing belief in the definition of an individual's biological identity through technology, thus disregarding what so many science-fiction movies have predicted (for instance "Welcome to Gattaca"), i.e. that scientists will soon be capable of simulating living organisms endowed with supposedly unique properties (Crettiez Xavier, Piazza Pierre, 2006). In this sense, a basically "technological" solution resorts to supposedly "natural" features to generate trust, although the latter, as we shall see, is purely based on convention.

On the other hand, the State can be construed as a last resort against self-proclaimed security "authorities". The State is the "*causal* guarantor of all legality. In that sense, it is the sole guarantor of the founding role of the Law as a norm-setting influence" (Legendre, 1992, p. 241). The so-called "trusted third-party", whose role is essential in PKI systems, is a market reality, but precisely because it is market-based, the third-party is felt to lack credibility and legitimacy because people fear it might be biased and self-serving. The quest for a guarantor must lead to a final, incontrovertible source of trust. In contemporary societies, only States can take on this mythical role. This in turn raises the question of constant border-transgression in a network-based society where none of the phenomena that call for regulation are restricted within national borders. The concept of security and the difficulty of implementing the concept in everyday life, show that the traditional guarantors, the bulwarks of legitimacy, no longer operate in a global society. Nor can the usual standardisation authorities sustain the necessary level of trust: what is needed is something than can recreate the mythology of trust, an entity that is seen to be independent of overly political bodies. Even freeware communities, such as Debian, are involved in this quest for the ultimate guarantor, free of any social or political impediments. In order to be recognised by the developer community, members take part in actual face-to-face meetings, so-called "key-signing parties" where each member can identify the other visually (i.e. the unique embodiment concept) and associate that visual authentification with the identity conferred by official ID cards and passports delivered by State authorities! By implementing this joint face-to-face and ID-based authentification process, the freeware communities bring together both physical and authority-based forms of guarantee, clearly stating that no garantee can be produced through an exclusively digital process.

It then becomes easier to understand why service companies offering security solutions are very reluctant to take on the role of "trusted third party", as our survey showed: the notion of trust refers to a more symbolic register, as the anthropologists would call it, which assigns roles to each player both within a cosmos and within a traditional hierarchy (transmission and inheritance are the basic conditions underlying guarantee and belief). Deciding who has "authority", either in the almost technical sense of certification authorities or in the almost disciplinary sense of who has the authority to decide within a given institution, mobilises a whole chain of delegation of trust and a body of beliefs, because no authority can be self-proclaimed, even though this contradicts the idea of technology as an all-powerful and overriding force. This in turn raises questions regarding justification, socio-political debate, the collective decision-making process leading to agreement, and all the other analytical processes that require a different kind of technical skill or know-how.

This same question arises when turning to more practical, everyday issues such as the electronic signature. Establishing the link between a function, a position in a management structure, a socially identifiable person and their signature would not raise any particular technical or symbolical problems if every correspondence was unique. In actual fact, all these dimensions interact in multifarious ways, both quality-wise and quantity-wise: one unique

person may have different functions, a same function may be shared by several people, signatures may de delegated, identical functions may correspond to different positions in the hierarchy or may vary within a same project, different signatures may be required according to status, documentation, project, etc… Each time, variable "attributes" have to be identified in coherent and very concrete and material terms, in a wealth of different situations. Every time, a decision has to be made regarding the principle underlying the correspondence between a given signature and one of the dimensions of a given entity, and this raises questions about the whole chain of authority and problems that cannot be solved simply by an all-embracing logical or mathematical modelisation of the system.

IV. SECURITY: THE NEVER-ENDING QUEST

All these technical, material and institutional conditions seem to be so complex and constrained that we can only surmise that they are designed to meet a pressing demand for greater security. Yet this is rarely the case.

A simple question such as "What is your attitude to security in your institution?" can elicit quite disturbing responses: "*The less I'm involved in security issues, the better"*, or "*I really don't know what to think, because I've never felt any real need for security*".

Such responses fly in the face of the mainstream idea that there is a growing demand for security. Yet most people seem to agree that present security provisions are *adequate*. How is it that our respondents appear to have so little interest in security issues?

Let us have a closer look at the ideas underlying this apparent lack of demand for security, as this has a significant bearing on how the market is organised. In this section, we shall give some insights on the ways people make do with security issues:

- considering it as always adequate;
- normalizing the deviant behaviours;
- fighting for deciphering the grammar of security procedures;
- trying to keep the motivation for applications that work and not for using the security system as such.

IV.1. Security levels always "adequate"

All security professionals, including firemen (Boullier and Chevrier, 2000) would probably agree: people only become interested in security once it is too late. In ordinary, everyday circumstances, people very quickly forget the lessons learned in times of crisis or the warnings of the professionals. Security is supposed to remain invisible and unobtrusive for the user, which means that people become accustomed to any given level of security. Habit is the major factor in this case, rather than any objective criteria. None of the respondents were able to formulate any clear scale of what constitutes a tolerable or intolerable security threat. None stated that they would consider this or that specific security measure unreasonable. Respondents tend to be very vocal when it comes to their independence, their individual free-

dom, but no one offered a definition of what they would consider offensive with regard to those basic principles for which academics are generally prepared to make a stand. Even the most critical tend to adjust and "make do". The main worry is to ensure the quickest possible "return to normal", because for most people, nothing is more challenging than having to constantly change their routine.

This kind of "satisfaction by default"(people are happy as long as they do not have to worry about it) poses a real marketing conundrum for those purveying security solutions: how to stimulate demand and make it more explicit. The kinds of public attitudes described should not be seen as implying the rejection or negative appraisal of security issues in general. But everyone involved in this field knows from experience that "too much security kills security", because overbearing security measures and a constant drive to challenge routines (which is what an effective security policy implies) generate so much "hassle" that people end up implementing avoidance tactics. A PKI development scheme cannot avoid generating some kind of demand, because a successful security scheme requires staff support, failing which any unjustified or unnecessarily annoying constraints are rapidly bypassed. Conversely, people are sometimes prepared to take on board new security practices, as we saw during the implementation of the French government's "Vigipirate plan" (public anti-terrorist awareness-raising measures) that no one seriously questioned. This is a facet that the security experts are familiar with and know how to exploit, because tragedies are the only time when they can introduce new rules in the hope of changing public behaviour. Providing of course these behavioural changes last long enough to become part of a routine and be forgotten.

IV.2. Deviance as normality

The idea of "adequate security" as the norm is justified on the part of the institutional operators in so far as any implementation of a security process goes hand in hand with the idea that individual behaviour must be placed under surveillance. Any security policy worth it metal has to implement its own "policing" mechanism. The question is no longer "who can do what?" but "who has done what?". Security implies sharing responsibilities and allocating blame. In this respect, it would be preferable to talk of a *juridical* security policy, as it consists in seeking to make words, objects and actions *assignable*. As Bruno Latour remarks: "We can indeed consider that the functional system that enables us to link, trace, assemble, relate, suture, and knit together that which enunciation by its very nature strives to distinguish, is part of that attachment that tradition designates as Law" (Latour, 2004, 35).

This is an aspect that often worries users and helps to reduce the demand for greater security. Any security policy implies a certain number of standards and rules that have to be complied with: "you have to lock the door", "you have to wear a badge", "you must sign such and such a document" etc.… But what happens when someone looses their badge? When something has been stolen because the door was left unlocked? When funding has been misallocated? Demand for security is inconceivable without a precise chain of responsibility. To put it more succinctly, implementing security means increasing the chances that the rules will be broken… a paradox with which lawyers are all too familiar. But it can become a hindrance to action in certain types of situations and contexts. With a view to greater rationalisation, a manager demanded that all the people under him transfer their computer passwords.

But this caused great disquiet in the department because people started to think: will he have access to my private data? Will he be able to check how long and how hard I have worked, etc?

In order to make the system more viable and tolerable, people are naturally reluctant to "grass" on their colleagues every time a rule gets broken. The system has a kind of in-built deviance tolerance. This was clearly illustrated in the "safe area" of one of the engineering schools we studied. In order to reduce the sound level (e.g. a door banging noisily every time someone opened it) or to avoid having to carry a badge every time someone left the area for a few minutes, someone decided to block the door open with a telephone directory or a curled up magazine… and the carefully protected safe area immediately became wide-open to all-comers. This was clearly a flagrant breach of the very security rules that originally warranted the setting up of the safe area. And yet this kind of behaviour will only be met at worst by grunts of disapproval, and at best by a few raised eyebrows and ironic smiles. In this case, deviance is tolerated. This kind of "make do" behavioural model (arrangement in the terms of Boltanski and Thévenot, 1991) in fact relies on a mutual deviance tolerance arrangement predicated on the principle: "you turn a blind eye to my failings and I'll disregard yours."

People are thus spared the need to justify their practices but the organisation gradually slips into a "my lips are sealed" mode where all kinds of occult influences can gain a foothold. This kind of mutual arrangement where security is always deemed to be "adequate", flies in the face of "institutionalised security" and any rational discussion of security policy. Mutual arrangements can also be explained by the fact that institutional security policies always aim for maximum security and can therefore easily be felt to be "over the top". The best thought-out security systems tend to allocate blame (and therefore punishment) to even the most seemingly benign kinds of security breach, precisely because this kind of transgression is seen to be the "weakest link" in the security chain. Making everyone accountable to everyone else can lead to an atmosphere of permanent suspicion, which most people will normally try to avoid by resorting to informal arrangements.

In this context, conflicts arise between:

- those who support an increasingly formal approach to security policies, on the basis that human behaviour remains the weakest link, thus generating further avoidance strategies,
- and those who recognise the need for an "adequate" level of security that can be implemented while making allowances for human foibles.

The security issue is thus constantly torn between these two approaches, and this kind of opposition is not conducive to greater commitment on the part of the actual players, whereas everyone knows that a security policy can only be effective if it can rely on total commitment of the people involved.

IV.3. Deciphering the grammar of security and the paradoxes of transparency

Any technical formatting applied to a specific institution resorts to a given vocabulary, to entities and categories that can take on very different meanings according to the people involved. The security levels, technical certification protocols, different authorities, etc., as descri-

bed by PKI system developers or even by the legal experts, are unlikely to be understood by ordinary users. Rendering the system totally transparent and thus making it unnecessary for the user to understand what is happening through automation, can prove to be counterproductive, or even legally indefensible, if the process is reduced to ritually delegating the user's signature to the application, which can lead to surprises when the expected guarantees prove to be totally inadequate.

This kind of "transparency" is ambivalent for the user in an overall cognitive ergonomics approach. Thus, "operational transparency" may, from a cognitive point of view, seem totally opaque for the user. A contradiction may therefore exist between an application's "transparency" and the user's comprehension of "what is happening". Particularly in the case of PKIs, the user's understanding of "what is happening" when he or she activates a security related function is the very basis of the security system's effectiveness.

The message asking the user for "permission to read" their private key whenever the decrypting function is activated, encapsulates this kind of dilemma:

"So when I installed the certificate on my workstation, I asked to be warned every time an application tries to access my private key. This is why this dialog box is displayed, [i.e. a pop-up window asking the user whether they authorize an application read-only access to his/her private key] and all I do is then press enter"*(operational manager)*

In this case, the chain of operations that the user has to carry out to implement the security function is *lengthened*, thus reducing the *operational* ergonomics qualities of the system. But at the same time, user security is *heightened* (the user knows that no one will read their private key without their knowing) thanks to *better cognitive ergonomics* (the user has a more thorough understanding of and greater control over what is happening). The head of computer services in one of the major companies we studied explains that they are currently considering three possible user configurations:

- no confirmation of private key reading by the application (which amounts to total user "transparency");
- a simple access confirmation by clicking on OK in a pop-up window;
- a PIN code password based access *confirmation* process, (for instance via a smart card).

The service company's technical manager has a more definitive position:

"If…I can simply access the reader, and can then "sign" any number of documents, that raises a major problem. … There are readers that work like that…All I need to do is identify myself to the reader, and then I can send any number of e-mails. This amounts to automation. There has been no physical authentification of my signature, and therefore no valid authentification. Without a subsequent approval of my signature, it has no validity."

This example shows that when it comes to security, "transparency" can actually prove *counterproductive*, not to mention the legality of the issue. We are facing here the same contradiction we already mentioned: naturalisation of security process is the only way to obtain use of the systems but this naturalisation is the surest way to lower the quality of the security chain.

PKI "transparency" is therefore not as simple as it seems: it implies compromises that can question the very core of the security system.

Unlike physical security, digital security cannot be easily apprehended and visualized by the ordinary user. The "grammar" of electronic security is difficult to grasp for a non-specialist, starting with an issue that lies at the very heart of Public Key Infrastructure, namely the assymetric key system: how can anyone who is not a specialist or who has not had chance to

learn about the system, understand that PKI security is based on a freely available "public" key that has to be combined with another confidential "private" key? How can one explain that the two keys are complementary but that this feature is only traceable one way (private key to public key) but not the other, and that the whole process is based on an unsolved mathematical problem involving large prime number factorisation? And if someone cannot grasp the basic concepts and tenets, how can they trust a system that they cannot apprehend via their senses[4]? Lack of comprehension leads to lack of trust in the system, and thus reluctance to take it on board and develop new uses for it.

Users in fact apply common thought processes and layman categories to security issues in order to construe a more coherent pattern of understanding for their behaviour. Trust and reputation are built on social practices and these can act as benchmarks for certain actions in an unknown and constantly changing universe. The perspective in this case is neither that of the engineers who design layer upon layer of security technology covering every possible case that their ontologies have identified, nor that of the lawyers who try to bring to bear the long-term historical perspective of the law. Only by adopting the user's point of view and by exploring the real or perceived institutional obstacles as seen by the user, can one redefine the essential issues. Only by basing ontologies on a redefinition of the limits of the institution and of the system (what is in and what is out), of the guarantor authorities and of the threats it is subjected to, can the true "grammar" of security be made accessible to all. This leads to "variable geometry ontology" (Latour, 1994), a major challenge for computer scientists in the field of security. Computer models based on ontologies have great difficulty in dealing with the constantly changing attributes for the entities that they are supposed to describe and the context-dependent properties for those entities. Only emergentist models can account for such processes and offer technical solutions, providing the loop includes all the so-called "social" parameters.

IV.4. Software applications as prime movers of security dissemination? Bridging the gap between transparency and understanding

A brief reminder of a truth that should be self-evident: nobody "uses a computer to generate security", and nor does anybody "set out deliberately to use PKI software". This type of functional distinction is meaningless in terms of user's activity, in cognitive ergonomics approaches (i.e. the distinction between "task" as prescribed by the division of labor and "activity" as a ongoing process oriented towards a socially meaningful goal). For the user, what is meaningful is the production and exchange of different levels of information, through different kinds of networks. The problem, when pretending to develop an all-embracing security system, is precisely to take into account all the different variables, different types of exchanges and types of information, within different types of networks!

This is why although the software applications are the "prime movers" in the PKI, they simply becomes for the user an underlying, invisible feature, or at best, visible in the shape of

4. This kind of issues has often been studied by anthropology of innovations, from the diffusion model school for instance.Rogers (1963) reports the failure of health programs in Peru that were unable to explain the power of microbes that cannot be seen by human eyes, although making things visible is a major pretention of modern scientific and naturalistic mind, against the animist, totemist and analogist ones (Descola, 2005).

a button. In terms of ergonomics, visual security level indicators can be integrated and accessed directly by the user, but the alternative should not simply be between the button that makes everything "invisible" and therefore impossible to check and understand on the one hand, and the parameter setting console that requires an advanced level of expertise. In an application, security level display bars should give easy access to a coherent visual representation of sets of complex data.

Visual displays should offer the security levels that are redefined by an ontology combining certain properties on the one hand and the levels of reliability of the recommending or guaranteeing authorities who have registered the signatures (for instance) on the other hand, as is the case in the freeware community. It is important to make visible the chain of delegation of trust thus produced, because trust only operates by delegation, and never by direct reference to the ultimate authority. Trust is typical of this contamination process (Tarde) that is best suited to understand the so-called "social" processes. It is more efficient to render visible the technical and human processes engaged in the certification of each link in the system, than to let everything rest on a hypothetical ultimate guarantor (as well as on the supposed cohesiveness of the whole structure).

V. BRINGING A "NON-STARTER" MARKET TO LIFE

The practices we have been able to observe and the experiences we have recorded have all been set against a background of "relative" security, far removed from the concept of security system "certification". This contradicts assertions about PKI as "the" ultimate solution to security and trust problems encountered in electronic networks. Slogans to this effect either reflect a purely technical determinism (the PKI concept and PKI technology as a guarantee of absolute security) or advertising hype (PKI: your security solution).

Problems relating to the sale of security solutions are particularly visible when it comes to offering clients a higher level of security, where the client-cum-user is identified by signatures and certificates based on a PKI.

– *"The other problem that banks face is how to charge their clients for improved security.*
– *And yet don't some banks charge for extra online services…?*
– *Yes, but they charge for special services, with a password login. So what should they do? Increase their basic charges to incorporate the extra security level? But then how will they be able to sell that to their customers? Can you imagine them having to explain that their services weren't that secure before they introduced the new charge?"*
(Service company)

The same problem of how to "sell" security is just as acute inside organisations themselves. The case of one organisation that we studied is particularly interesting in this respect. The organisation had developed a PKI for in-house use, which did not require a qualified signature protocol. They deliberately chose not to go down that road and to simply aim for *improved* security, which means that their certificates had no legal basis.

The problem arises when the company then wants to "sell" the services generated by that infrastructure, and particularly non-certified signature services, to other users.

These examples show the problems related to security sales strategies:

- Home banking clients who are persuaded to "buy" a first generation of supposedly "secure" online services and who are then offered a second generation of services on the basis that the initial services were not all that secure after all…
- Potential users of a "*high risk*" organisation who are offered a signature system on the basis that it will make communications more secure, and who will soon be informed that the system was not in fact a bona fide signature system… and that a new one will now be installed!

It is little wonder, seen from this angle, that the market has not yet "taken off" and that organisations find it so difficult to deploy security systems. Is it really surprising that security systems and products sales people have difficulty persuading potential clients to sign the order form? Or that computer systems managers or security officers find it difficult to persuade the MD to invest in the latest security system? Or that project managers find potential users reluctant to take part in a trial run?

Building up demand on the basis of rational argument is made more difficult by the very fact that the actual risk is *impossible to quantify objectively*, as by definition, it has not materialised and is impossible to prevent with 100 % certainty.

This dilemma, which lies at the heart of risk economics, can be partially resolved by risk identification and risk level assessments applied to each particular organisation. But this is a costly process and many organisations are reluctant to spend money on a risk audit which will necessarily reveal organisational weaknesses. Again, barring special circumstances such as a recent security failure, it proves very difficult to "sell" security loophole diagnosis.

Moreover, any potential software solution has to be seen as part of a general security strategy. The actual value of any potential security hardware and software investment will depend first and foremost on the activities that the company develops on the basis of these goods and services, and on the organisation's specific security cultures and practices.

But the economic value of a signature and of any underlying PKI investment, is even more uncertain if it does not have a legally valid, qualified status. A technical PKI solution is supposed to "guarantee" computer network security within the organisation concerned – on the basis of the value of the signatures, and of the robustness and inviolability of the asymmetric prime number key algorithm. Yet according to one PKI implementation manager: "The computer engineers tell us that *electronic signatures will be unquestionable!* Of course, not many people are likely to question the PKI algorithm. [But] they will *question the way it is implemented*" (Adjole).

Three types of criteria that set the economic value of a secure signature system can therefore be identified:

- the technical, organisational and institutional *resources* that the security operator uses to back up the relative robustness of the security system they are implementing, and of the signatures derived from it;
- the extent of the *liability,* and particularly the financial liability shouldered by the issuer for each security level, according to the different tariffs and different uses allowed for each type of signature;
- the degree of *trust* awarded to a particular institution, based on a subjective assessment of the credibility of the assurances made by the institution as to what it can deliver.

These three sources of uncertainty regarding the value of the signature are the main reason why potential users are so *undecided*:

- the degree of trust awarded to the operator's assurances,

- the "expectations" that the client has regarding what the system will offer in the event of an actual security breach and what proof they will have of the system's effectiveness,
- who will actually be responsible for supplying that proof.

In the organisation mentioned earlier, not only did they decide not to go for "legal" recognition of their signature, but they also did not want to attach any liability, at least initially.

VI. CREATING A MARKET FOR A PUBLIC GOOD?

The development of online business activities using Internet, Extranet or Intranet-based electronic networks is generating an urgent need for the electronic equivalent to the "paper" signature. There is definitely a macro-economic demand for the digital signature as a source of further economic development in the knowledge-based and network-based society, but market forces alone do not seem to be capable of generating commercially viable products in this area on a long-term basis.

Economists would call this kind of product, which meets an overall demand but cannot be supplied by private operators under normal supply and demand terms, a *public good* or *collective good*. So what if authentification-signing was in fact a public good, just like national security and justice, to be organised and funded, at least in part, by the State? This would satisfy what we described earlier as the users' quest for a universal guarantor, a quest that is presently hindering online business development. It would also acknowledge the fact that a PKI electronic signature system is first and foremost an *institutional* process, and not simply a technological or organisational one. Finally, it would recognise that exchange systems only work effectively in a networking context. The *usage value* of a signature, namely what a user gets out of using it, is closely related to what use other people make of the system. This is what economists call the "external benefits" or external effects of the network.

It is worth recalling at this point what the electronic signature services manager told us: "*I personally would have nothing against the State carrying out public certificate authentification. As my business revolves round digital signature-related services it would actually provide good leverage for my business. At the moment, nobody knows where they stand with the certification authorities market, and no one wants to put any money into it.*"

Nevertheless, this does not mean that things are simple. PKI and electronic signature technologies, markets, clients and institutions are not only still largely fuzzy, undefined entities, but they are also mutually dependent and their evolution is closely interlinked.

There are two main arguments in favour of developing a "flexible" legislative and institutional framework, sufficiently sustainable for it to be adapted when the need arises. For a start, electronic signature technologies and markets are still in the emerging phase from an industrial point of view and therefore require enough leeway to be able to explore various technical and economic combinations. It is impossible to completely regulate a field of activity while the contours are still blurred. Secondly, electronic security goods and services are very specific in nature: effective security is a relative and imperfect concept. Any regulations that would purport to sustain "absolute" security would necessarily be inapplicable and would be likely to inhibit any further development in the area rather than the opposite.

Qualified imperfection engineering, based on everyday usage, and the *"formatting" of users*, so they become accustomed to the common "grammar" of security, seem to be two major and often neglected keys to the successful implementation and management of PKI systems and their widespread acceptance within organisations.

Methodology

The observations described above were carried out as part of an RNRT I-Care project (Trust-generating infrastructure on internet and mobile networks), that took place between September 2001 and September 2003, with the following partners: Thalès, the ENST, Eurecom, The Alès Mine Engineering School, Mr T. Piette-Coudol, a lawyer, CEA-LETI, and Francert. We had to take into account the difficulties that we anticipated in gaining access to ordinary businesses equipped with standard PKI systems. We therefore used a roundabout approach, subdividing the object of our research into two sub-strands: this allowed us detail the social and technical specifications of asymmetrical PKI-based security systems. This kind of methodological difficulty was in itself symptomatic as it only confirmed the initial premise, i.e. that the techniques exist but they are rejected by the market. Security is indeed a problem, but users have no desire to take it seriously.

1/ We therefore conducted observations first of all on sites that were likely to be equipped with the systems, i.e. two engineering schools, and attempted to determine the organizational issues underlying current practices which could guide future specifications for development or deployment strategies. The observations were based on 21 in-depth interviews with different people working in academic, technical and administrative positions. All the interviews were noted verbatim and completed with detailed observations on all the research objects and the ecological framework of the activity. In accordance with ANT principles, we were careful not to give a finite pre-conceived definition of security, and focused on the contrary on the actors' capacity to establish links between different meanings, issues and actors in their practices and discourses. The very fact of putting forward the need for security and acting according to different principles does not constitute in our eyes either a statement of faith or a contradiction in terms, but represents different ways of building a shared world acceptable to all. In this perspective, we concentrated on the benefits of taking into account actual practice, without going so far as to consider that actors can achieve a form of total control over their own utterances. They are constantly being overtaken by their own practices, and this is why security still manages to function! Our constant guiding principle was therefore to take every element in their discourse very seriously, even when it appeared to be contradictory.

2/ Getting feedback from real users of actual commercial systems proved more challenging, because as with any network effect, the first user is always at a disadvantage and has to be more proactive in the sense that his/her partners are not using the same system. We therefore opted to focus on very different user sites, trying to aim for a diversity of partners over time (i.e. 11 interviews repeated with a 12 month interval) from:

- a major company, first during the pilot phase then «for real», (2x8 interviews plus observation)

- a freeware network –Debian- with frequent users (2x2 interviews plus active observation)
- a small business selling a security solution (2x1 interview)

The emergence of the need for security and its actual manifestation were radically different in each of the three cases. In total: 21 interviews on sites with incipient security policies.

Manuscrit reçu le 17 octobre 2006
Accepté le 11 juillet 2007

REFERENCES

[1] AKRICH (M.), « Les objets techniques et leurs utilisateurs. De la conception à l'utilisation », in B. Conein, N. Dodier, L. Thevenot, Les objets dans l'action, Raisons pratiques, Paris, *Éditions de l'École des Hautes Études en Sciences Sociales*, pp. 35-57, 1993.

[2] AKRICH (M.), « Les utilisateurs, acteurs de l'innovation », *Éducation permanente*, n°134, pp. 79-89, 1998.

[3] AUDREN (F.), BOULLIER (D.), JOLLIVET(P.), « L'institution de la sécurité ou comment s'en désintéresser », rapport RNRT, Icare (Infrastructure de confiance sur des réseaux Internet et mobiles), 2002.

[4] BOLTANSKI (L.), THEVENOT (L.), « De la justification. Les économies de la grandeur », *Gallimard* (NRF), Paris, 1991.

[5] BOULLIER (D.), CHEVRIER (S.), « Grammaire de l'urgence : les sapeurs-pompiers, experts du risque », *Les Cahiers de la Sécurité Intérieure*, n° 22, 1995.

[6] BOULLIER (D.), CHEVRIER (S.), « Les sapeurs-pompiers : des soldats du feu aux techniciens du risque », PUF, Paris, 2000.

[7] BOULLIER (D.), « La vidéosurveillance à la RATP : un maillon controversé de la chaîne de production de sécurité », *Les Cahiers de la Sécurité Intérieure*, n°21, 1995.

[8] BOULLIER (D.), « Les conventions pour une appropriation durable des TIC. Utiliser un ordinateur et conduire une voiture », *Sociologie du Travail*, n°3, pp. 369-387, 2001.

[9] BOULLIER (D.), « Les études d'usages : entre normalisation et rhétorique », *Annales des Télécommunications*, **57**, n°3-4, pp. 190-209, 2002.

[10] CALLON (M.) (ed), "The Laws of the Markets", Basil Blackwell, *Oxford*, 1998.

[11] CALLON (M.), « Sociologie de l'acteur réseau », in N. SMELSER, P. BALTES (dir.), *International Encyclopedia of the Social and Behavioral Sciences*, Oxford, UK, Pergamon, pp. 62-66, 2001.

[12] CRETTIEZ (X.), PIAZZA (P.) (dir.), « Du papier à la biométrie. Identifier les individus », Paris, *Les presses de Sciences Po*, 2006.

[13] DELEUZE, Gilles, « Post-scriptum sur les sociétés de contrôle », *Pourparlers*, Paris, Éditions de minuit, 1990, p. 240-247.

[14] DESCOLA (P.), « Par-delà nature et culture », *Gallimard*, Paris, 2005.

[15] EYMARD-DUVERNAY (F.), FAVEREAU (O.), ORLEAN (A.) SALAIS (R.), THEVENOT (L.), « L'économie des conventions ou le temps de la réunification dans les sciences sociales », Problèmes économiques, *La Documentation française*, n° 2838, 2004.

[16] JOLLIVET (P.) (avec P. DIEUAIDE), « TIC, Artéfacts et Redéploiement de la Norme de Consommation : Que peut nous apprendre le modèle «Hacker»? », *actes du colloque Conférence SASE 2003* (Society for the Advancement of Socio-Economics), Aix en Provence, 2003.

[17] JOLLIVET (P.), « Les NTIC et l'affirmation du travail coopératif réticulaire », in Le Capitalisme Cognitif, C. AZAÏS, A. CORSANI et P. DIEUAIDE (eds), Paris, *L'Harmattan*, 2001.

[18] LATOUR (B.), « Une sociologie sans objet ? Remarques sur l'interobjectivité », *Sociologie du travail*, n° 4, pp. 587-607, 1994.

[19] LATOUR (B.) "Re-assembling the social. An introduction to actor-network Theory", *Oxford University Press*, Oxford, 2005. Trad. Française: « Changer de société. Refaire de la sociologie », *La Découverte*, Paris, 2006.

[20] LATOUR (B.), "Aramis or the Love of Technology", *Harvard University Press*, Cambridge, Mass. (translation by Catherine Porter), 1996.

[21] LATOUR (B.), « Le Prince : machines et machinations », *Futur antérieur*, n°3, pp. 35-62 1990.
[22] LATOUR (B.), « Note brève sur l'écologie du droit saisie comme énonciation », *Cosmopolitiques*, n°8 (Pratiques cosmopolitiques du droit), pp. 34-40, 2004.
[23] LEGENDRE (P.), « Les enfants du Texte. Étude sur la fonction parentale des États », Paris, Fayard, 1992.
[24] LEGENDRE (P.), « L'empire de la vérité. Introduction aux espaces dogmatiques industriels », *Fayard*, Paris, 1983.
[25] LIVET (P.), « La communauté virtuelle », *Editions de l'éclat*, Combas, 1994.
[26] ORLEAN (A). (dir.), « Analyse économique des conventions », PUF, Paris, 1994.
[27] RASMUSSEN (J.), "Information Processing and Human-Machine Interaction", North Holland, 1986.
[28] REASON (J.), "Human error", *Cambridge University Press*, 1990. Trad. Franç : « L'erreur humaine », PUF, 1993.
[29] ROGERS (E. M.), "Diffusion of Innovations", *Free Press*, New-York, 1983. (1st edition: 1963).