An Outlook to Security and Trust in Internet Communications

Robert Logozar
Polytechnic of Varazdin
J. Krizanica 33, HR-42000 Varazdin
robert.logozar@velv.hr

Petra Koruga
Faculty of Organization and Informatics, University of Zagreb, Pavlinska 2, HR-42000 Varazdin
petra.koruga@foi.hr

Abstract*. In this paper we address the key questions on security and trust in communication over Internet, in a synthetic approach aimed to connect the technological and human aspects of the subject. As first, a brief outline of the Internet security technology is given. This serves as a ground for the exposition of the general security concepts and principles, its pillars and threats. The achieved security provides the basis for building of the user trust. The trust is then proportional to the user's perception of the achieved security level. The omnipresence of Internet in all human activities today, including financial transactions, e-commerce, trade, auctioning, and other, proves by itself that the trust of majority of online users is won. Legal support and especially the mechanisms of the user protection initiated by the service providers are improving, as well as the general efforts to educate the online community.*

Key words: *Internet security infrastructure, security perception, technology-related and human-related aspects, trust, security statistics.*

1 Introduction

The modern era is characterized by a widespread use of many different communication systems. Among the most complex are computer networks, which have grown globally and locally, occupying the world-large scale and penetrating the inner organizational structures. They serve extremely large number of separate users, connecting them to local or universal communities. The computer networks are known of their diversification, wide variety of protocols in use and huge quantities of information transmitted over them. The data transfer relies more and more on the network facilities, which become an integral part of the computing infrastructure. Technically speaking, they present information channels with many different security threats. Along with the benefits of the intensive networking, the need for thorough security solutions emerges as more and more crucial. Via Internet we can acquire the newest antivirus program or a patch for our operation system, but in the same time we can expose our computer to malicious attacks.

The technical complexity of such a channel often contributes to mystification, misunderstanding, exaggeration or underestimation of the security issues. As we approach the third decade of widespread use of Internet in all spheres of life, we must say that the trust of majority of users is improving. It is based on both, the widespread trends, and the personal experience. Still, when we ask simple questions, such as:

- Can Internet be securely used for human communication?
- Can privacy and identity of Internet users be protected?

broad users may be perplexed. On one hand the answers should be affirmative, judging by the facts on enormous number of successful delicate online transactions happening as we speak. On the other hand, the reports of Internet frauds can cause disbelief and mistrust, and raise the questions about the involved risks.

The aim of this paper is to describe the existing foundations of the Internet security, and to help the readers to answer the above questions. We start with a low level description of the security technology infrastructure, aimed to wider audience. Upon that we build the interdisciplinary approach which shows that in a complex system aimed for human communication and interaction, all participants are important. Every single computer counts! It may be a brick in the global security wall, a hole in it, or a source of unintentional or even intentional threats and danger. In other words, the technological solutions present a basis that must be provided, but by itself cannot be considered as a *full security foundation*. The human influence is unavoidable and it calls for a *synthetic* and *interdisciplinary* approach. In other words it is depicted as *multilateral* or *multidimensional*, showing clearly that the security field outgrows mere technological domain. It is also a psychological, ethical, economical, legal and political issue [1,2].

With respect to the above, our intention is to provide a better insight into technological security aspects to the readers with social and humanistic background, and to outline the importance of human and social aspects to the technical audience. Throughout the discussion, we draw a line that is common to all communication systems, making evident that Internet is not at all that specific. Most of the problems do not originate from Internet. They have just emerged here as more blatant, due to the Internet's vast potentials. And like any other media, Internet has its advantages and its limitations. Such thinking will lead us to the more general treatment of the security issues.

2 A crash review of Internet security infrastructure

Here we briefly interpret some technical features of Internet and relate them to security aspects of communication. Although it may seem as futile to even try to bring this large subject in only one section, this review goes along with our thesis that only an educated Internet user is a risk-aware and, henceforth, a safer Internet user.

2.1 The Internet model and OSI model

Technically, Internet is described by the *TCP/IP model* (also called *Internet model*), or *Internet Protocol Suite*. The first name is after the Internet two most important protocols. The model can be divided into four *abstraction layers*, which are outlined from the top down in Table 1 [3,4].

Table 1. TCP/IP model, or Internet Protocol Suite.

	Layer (protocols)
4	Application Layer (FTP, HTTP, SMTP, SSH, SSL, TLS, …)
3	Transport or Host-to-Host Layer (TCP, UDP,…)
2	Internet or Inter(Network) Layer (IP, IPv6, IPsec, …)
1	Link, or Host-to-Network (ARP, PPP, DSL, ISDN, FDDI, …)

A *communication protocol*, or shortly *protocol*, is a procedure that precisely describes how the communication is to be done. Typical protocols corresponding to each of the layers above are listed within the parenthesis. The Transport and Internet layers, and the corresponding basic TCP (Transmission Control Protocol) and IP (Internet Protocol), present the core of the Internet as we know it. There are a few more protocols in these two layers, as there are several more protocols in the top-most application, and the lowest link, layer.

In fact, the layering was not part of the original TCP/IP specifications. The concept was introduced by the OSI (Open Systems Interconnection) model. The OSI model remained mostly within theoretical realms, but its good solutions largely influenced the way of analyzing and developing of the computer networks, including the TCP/IP model. So, although strict comparisons are not fully justified, most authors try to provide some mapping between the two models. The seven layers of the OSI model and their "rough" relation to the abstract layers of the TCP/IP model are shown in Table 2.

It was mainly due to the OSI model that the link layer of the Internet model is usually divided into two sublayers: the *Data Link* and *Physical Layers* (or *Network Interface* and *Hardware Layers*), bringing the number of Internet layers to 5. In short, starting from the bottom, the Internet Link (Host-to-Network) Layer corresponds to OSI layers number 1 and 2 – the *Physical*, and *Data Link Layer*. The Internet Layer corresponds to OSI layer 3 – the Network Layer. The Transport (Host-to-Host) Layer is mapped to the OSI layer 4 with the same name (though their precise definitions defer), and also partly to the OSI layer 5. The application layer roughly corresponds to the OSI layers 5 – 7: Session, Presentation and Application layers. In the OSI Presentation Layer the encryption was predicted, allowing the syntax of the application layer to be independent from the selected security solutions, and also from the functions of other lower layers (confer 2.4).

Now we can follow the "layer stacks" of the two models, and briefly sketch the layers' functions.

The *Physical Layer* specifies electrical properties of the networking devices and their interfaces to the transmission media (copper lines, optical fibers, radio frequency electromagnetic waves), through which it sends the bits of data. It defines the connectors' pinouts, voltages, clock-rates and other technical details of the network hubs, repeaters, network interface cards, routers, and other devices.

Table 2. The abstract layers of the TCP/IP and OSI models and their rough relation.

TCP/IP model		***OSI model***
5. Application Layer		7. Application Layer
		6. Presentation Layer
		5. Session Layer
4. Transport Layer		4.Transport Layer
3. Internet Layer		3. Network Layer
Link Layer	2. Data Link L.	2.Data Link Layer
	1. Physical L.	1. Physical Layer

The *Data Link Layer* functionality, as is valid for every higher layer, is based on the services of the lower, physical layer. The data link layer provides transmission of digital data organized in *frames*, between the hosts on the same network (LAN, WAN, confer 2.6), from one end of the transmission media to the other. This layer provides a service interface to the network layer above it, by checking and correcting the transmission errors. It also regulates the data flow on the basis of physical addressing, taking into account the capacities and speeds of the sending and receiving devices.

The *Internet* (or *Network Layer*) provides the transfer of *data packets* from a source host to the destination host specified by an IP address, within the same network, or on different networks (the latter is also known as *internetworking*). This is done through the process of *packet routing* in which the packets are sent to the next network node (realized by the functionality of a *router* device) on the patch to the final destination. The network and Internet topology must be known in order to ensure the packet transport via routes which avoid congested communication lines and routers. This is the lowest layer that provides the End-to-End connectivity. Its functionality is today provided by IP.

The *Transport Layer* uses services of the network layer to ensure the End-to-End transfer of the messages from a process on a source computer to a process on a destination computer. This layer assures flow control, congestion control, and application addressing (port numbers). It provides the necessary abstraction level for the work of application software in the layer above, assuring that it is independent from the lower layers. The main protocols of the layer are TCP and UDP. The TCP provides the so called *connection-oriented* data transmission, and UDP provides the *connectionless* transmission of *datagrams*.

The topmost *Application Layer* is used by applications for specific network communication tasks. The layer presents the higher-level protocols: FTP, SMTP, HTTP,… For us, the interesting protocols are also the security-providing ones, like SSH, SSL, TLS, which will be specifically mentioned in 2.4. Generally, the application data is formatted and coded according to these protocols, and is then *encapsulated* into the protocols of the lower transport layer. They in turn use the services of the protocols which are lower in the layer stack.

2.2 The lack of security in the basic Internet layers

Internet misses a true and convincing security concept in its fundamental Internet and Transport layers represented by the corresponding IP and TCP protocols. The unbelievable historical success of Internet is based on the fact that it is relatively simple, fully open and decentralized network, not belonging to anyone. Its reach is global, but there is no global control of its functioning. The comparison to the openness of human society is striking. There is no true global security policy in the human society either (at least not today). The control of security measures is implementable only locally. On the global scale the situation varies and uncertainty prevails.

As was already stated in 2.1, IP (Internet Protocol) deals with data packets, self-contained, independent chunks of information bearing the IP address, and the associated mechanism of *packet switching*. This basic concept provides much of the functionality of Internet communications, like the optimal use of resources, great flexibility and low cost[*], but it also introduces additional security risks. As opposed to the *circuit switching* found in telephone connection, the traveling path of information on Internet is more arbitrary and not at all certain. Since the information is in digital, "electronic", form, it is furthermore prone to low-cost and easy-to-be-done subversions and attacks. Namely, with today's digital technology, the electronic digital data are not only the most easily stored, transferred, received, and protected from noise—comparing to all other forms of data presentation and physical realization, like those written on paper, or analog signals modulated in radio waves—but are also the most easily copied, altered, multiplied, forged, etc. Because of that, the proper protection of data and implementation of security mechanisms is of utmost importance (confer also chapter 3).

The examples of the low-cost threats are: "password sniffing" (searching for non-encrypted passwords by programs installed e.g. on the servers placed on the network backbones, "IP spoofing" (finding the IP address information within the packets IDs and using them maliciously), password stealing (e.g. by Trojan horses thrown into the system), etc. All these attacks can be performed in every node of the network that is traversed by data packets of a message.

Internet is known to be open both horizontally, for free spreading of the network, and also vertically, meaning that new protocols can be added. But the vertical openness could require changes in the infrastructure, which is hard and expensive to implement. Also it could present a source of incompatibility and restrictions for its horizontal openness.

The basic TCP/IP architecture of Internet is non-cryptic in its nature. This immediately allows for the loss of secrecy and loss of integrity, because of the attacks performed in any of the Internet layers. The usual, unsecured Internet services, such as electronic mail and file transfer, are unprotected from such attacks. Yet another common problem, unsolvable by the original Internet infrastructure, is the lack of authentication[†]. Without it, any higher forms of business communication cannot be realized.

Both of the problems can be mended by adequate use of cryptography mechanisms. The security defects were partly remedied in the mid 1990s by introduction of the End-to-End security protocols IPsec and IPv6. They ensure security mechanisms in the internet layer by authenticating and encrypting each IP packet. The idea was to alleviate the burden of the security implementation from the application software. However, the need for implementing and maintaining the dedicated software for this protocol on every remote computer, resulted that the security solutions in the application layer prevailed (see 2.4).

2.3 Cryptographic solutions

To make further discussion clearer, we shall briefly outline the basic cryptographic concepts (for more details see e.g. [5, 6]). There are two basic cryptographic systems in use: *symmetric cryptosystem* with *secret key*s, and *asymmetric cryptosystem* with *private* and *public keys*.

Symmetric cryptosystem was used in DES (Data Encryption Standard) a former American standard

[*] The packets bearing their ID numbers are transferred independently from each other, through different nodes and via different paths, enabling better overall usage of the available bandwidth.

[†] Authentication is the act of verifying the genuineness of an entity, i.e. the security process of establishing that the entity is what it claims to be, and that it can act as a known subject (person, process, computer, etc.). Only after the authentication, the *authorization* should be done. It is the process of verification that a known subject is allowed to perform certain actions and access certain resources.

from 1977, which was replaced by Triple DES in the late 1990s. In early 2000s AES (Advanced Encryption Standard) superseded DES and Triple DES, with its longer 128-bit code blocks and longer keys (128, 192, and 256 bits). A disadvantage of the system is that a safe channel must be used for the distribution of secret keys. Though the need for the extra safe channel can be regarded as a technical shortcoming, by establishing it between a known and certified sender and recipient, the problem of authentication of the communicators can be simultaneously solved.

Asymmetric cryptosystem eliminates the need for another safe channel by introducing a pair of keys, consisting of the *private key* (to be kept secret by a sendee) and the *public key* (to be disseminated to possible senders). The sendee (recipient) who wants to receive an encrypted message distributes his or her public key to the other side(s). The other side, the sender, uses it for encryption of the message to be sent back to the sendee. There's no fear that the message will be understood by any third side. The method ensures that decryption cannot be done without having the sendee's private key, which is irretrievable from the public key. Thus, only the sendee who owns the private key will be able to decrypt the message. The RSA system, named after its inventors (Rivest, Shamir, Adleman), is such an asymmetric system. Because there is no need for additional safe channel, this is an ideal solution for Internet, except that it requires much higher computing resources than the symmetric cryptography.

However, without having a safe channel (which would a priori assume the proper authentication of the communicating sides), an intruder can take someone else's, or generally false identity, and abuse it. So, in this case the need for a proper authentication emerges as crucial. The problem is solved by the introduction of Trusted Third Parties (TTP), which take over the distribution of public keys (other common abbreviations in use are: PKM – Public-Key Manager, PKDC – Public-Key Distribution Center). Then by the use of secure protocols (new protocols which include the cryptographic mechanisms — see the next section), a proper authentication is ensured, as well as that both sides have each other's public keys. With the proper authentication, the use of asymmetric cryptography simulates the possession of a safe channel.

The secure protocols and mechanisms use both cryptographic systems in order to ensure optimal results. Since the asymmetric cryptography is about two orders of magnitude (or even more) slower than the symmetric one, it is used only for the crucial parts of communication: for the authentication and for the encryption of the secret symmetric keys. After the symmetric keys are exchanged, the rest of the communication is protected by much faster symmetric encryption.

In the *digital envelope* data itself are encrypted symmetrically, while the asymmetric cryptography (simulating the safe channel) is used for transmission of the symmetric key only. Thus much greater speed of secure communication is achieved. Digital envelope ensures data secrecy, but not data integrity. Namely, although information remains secret to an intruder, it can be illicitly damaged or altered.

Digital signature solves the problem of the message integrity by calculating the hash function or *message digest* out of it, and then applying the asymmetric encryption to the digest. Both, the encrypted digest and the original message are sent. If the message is changed, the recipient will know it by comparing the original digest (after decrypting it), and the newly calculated digest from the received message. Only if the two digests match, the message is genuine. The mechanisms of digital signature and digital envelope can be combined together to provide joint secrecy and integrity. If public keys were distributed properly, as pointed out before, the digital signature ensures the authenticity, secrecy and integrity. Usually by the name of digital signature all these security mechanisms are assumed.

2.4 End-to-End security — cryptography in the application layer

The simplest way of introducing the security on Internet and leaving the lower layers of the TCP/IP model untouched, is to implement it in the highest, application layer, by means of cryptography. This is known as ***End-to-End* (*EtE*) *security in the application layer***. Thus, although attacks in the lower layers are not prevented, they are made futile with respect to many security aspects. This can be interpreted as introduction of a new, Security Layer., in our case based on the SSL protocol[‡], as illustrated in Table 3.

Table 3. The introduction of Security Layer [4].

Layer	
6	Application (HTTP)
5	Security (SSL)
4	Transport (TCP)
3	Network (IP)
2	Data link (PPP)
1	Physical (DSL, ADSL, cable TV)

This idea and the EtE security concept is implemented in several protocols aimed for different applications. These are:

- SSL (Secure Socket Layer) [7], already mentioned above, and now being upgraded by the newer:
- TLS (Transport Layer Security) [8];
- HTTPS (Hyper Text Transfer Protocol Secure) that is simply the usual HTTP over SSL or TLS;
- Secure Shell [9],
- PEM (Privacy Enhanced Mail), now replaced with:
- S/MIME (Secure/Multipurpose Internet Mail Extensions);
- PGP (Pretty Good Privacy) data encryption and decryption software;

[‡] The SSL was invented as a secure protocol within the Netscape suite of network applications in mid 1990s.

- GnuPG (GNU Privacy Guard) free cryptographic software;
- Etc.

E.g. S/MIME ensures the secrecy of e-mail communication over a non-secure network by the use of secret keys and symmetric cryptosystem, under the assumption that the local Internet servers are secure for key handling. PGP uses the asymmetric encryption with public keys for the critical data, thus removing the burden of keeping secret keys on servers (they are kept on the client's computers).

These were all examples of how cryptography and EtE security could successfully protect user data from the first two threats: loss of secrecy and loss of integrity.

2.5 The achieved security level

The way of measuring the achieved technical security level is by finding its *intrusion work* W_I. It is the computational work needed by an adversary to breach the applied security and undermine the system that can be expressed as:

$$W_I = P_{Cmp} \times t \ . \qquad (1)$$

Here P_{Cmp} is the computing power or speed of the intruder's computer expressed in some suitable manner, and t is the time spent on breaking the security by brute force. It is implied that the intruder is using the most efficient algorithms known. In the early 1990s, when the computers had the processor power of the order of 10 – 100 MIPS roughly (1 MIPS = 1Mega Instructions Per Second), the W_I used to be expressed in tens and hundreds of MIPS×Years. Because of the growing computational power of the computers, this intrusion work is not impressive any for a long time now. The computing power of the commonly accessible computers today has grown immensely, and even the ways (benchmark tests) of measuring it have changed in order to more accurately represent the performance of a computer system as a whole (e.g. SPECint, SPECfp). Roughly we can say that the computing power grew for the factor of 10^3 – 10^4, so that the intrusion work should be enlarged for the same factor.

The intrusion work W_I needed to break a key is rising greatly with the key length. For symmetric cryptography the rise is close to exponential, and the chances to break it by "brute force", i.e. by systematically trying all the possible keys, are extremely low. There are no reports of successful cracks by now.

The asymmetric cryptography requires longer keys for the same level of security than the symmetric one (roughly by the factor of ten, with the tendency of even bigger factors for bigger key lengths (e.g. confer [6]). In the combined systems, with the use of RSA for asymmetrical encryption of the symmetric key, the RSA is considered as the weakest link, most vulnerable to attacks. Though there are reports on cracking down the RSA system by the use of abundant computing resources and in cases of shorter key lengths [10], [11], the only credibly reported breaking of the RSA and particularly the PGP, as one of the most popular hybrid systems using it, was done by enormous computer and organizational power.

Anyhow, longer and longer keys, and improved algorithms are in use to ensure against more elaborate and sophisticated methods of attacks. Back in 1996 the symmetric keys of at least 75 bit length were advised, with suggestions to enlarge them to 90 to compensate for the rise of computer power. After the standard DES was replaced with Triple DES, and nowadays with AES, the key length is at least 128 bit for standard applications, and 256 bit for critical ones.

As for the asymmetric encrypting, 1024 bit and longer keys are not rare any more. Only a few years ago such cryptography was treated as a high-tech product strictly forbidden for export from the U.S.A. The high-level security needed in banks requires asymmetric keys of 2048 bits and even longer [12]. The safer solutions will require more computational resources and will be more expensive.

As a conclusion, by taking the key appropriate to the security demands of an application, the intrusion work can be designed high enough to make the security attacks not worth the effort [13]. In other words, if the cryptographic EtE security concept requires several months or even years of computing to be breached by brute force, than by ensuring a simple policy of changing the keys on a regular basis, together with other mechanisms of recognizing and stopping such attacks, we can make them futile.

Without going into further details, the simple conclusion follows: *in the context of building the communication trust, the use of cryptography must be made completely consistent and without exception.* Furthermore, the cryptography should be standardized and regulated more consistently, which is generally not the case. Poor cryptography was often put in large software packages [14], perhaps under the pressure of restrictive export regulations. As a final result, non-secure products could appear on the market, justifyingly adding to the users' loss of trust.

2.6 Intranets, firewalls and local security

As opposed to the global uncertainty of Internet, the Intranets, and generally LANs and WANs (Local Area Networks and Wide Area Networks) present the proprietary networks in which security polices can be established and enforced rigorously. Here the security on the **technical level** can be made highly predictable. The general defects of Internet can be, if not completely mended, at least kept under control. The intranet is interesting because it can use the standard Internet infrastructure (protocols) and applications, while enabling the full supervision of all the servers and clients within the localized network. Besides that, intranets can use other specialized protocols (like X.25) and networking solutions that can highly improve security (EBICS, SWIFT).

For intranet and other private networks, the basic security principle of connecting them to the "wilderness of Internet", is of doing it **only** via a strictly controlled protecting system, called *firewall*. The firewall is a hardware or software component, or combination of both, used to control the communication between different segments of network, specifically between the intranets and Internet, on the basis of set rules and policies. Mostly, the firewalls are set to control the traffic from some insecure and hard-to-control parts of the network, like Internet, to the local secure networks, or home computers. They should protect the "inner side" from the unauthorized accesses and threats from the "outer side", while allowing the desired and approved data transfer. Also, they should restrict the transfer of the secret data from inside to the outside world.

The firewall can be organized as one or more of the following:

i. *Packet filter*, which filters out the packets with respect to their departing and arriving IP addresses, and requested TCP ports (services). The filtering is done according to the list specifying the addresses and services which are forbidden, those which are allowed, and the rules of actions for the rest of the packets.

ii. *Application Layer Firewall*, which acts through the application software by controlling the IP packets coming to particular applications, like Web browsers, FTP clients, etc.

iii. *Firewall on Proxy Servers* acts similarly to the application layer firewall, but since they are organized as servers, either on separate computers or as software, they offer their clients additional level of security.

iv. *Firewall with Network Address Translation* (NAT) mechanism protects the computers behind itself by hiding their true IP address. This is usually combined with the standard role of the NAT (the enlargement of the number of IP addresses within local networks).

The firewall must be complemented with the intrusion detection system (IDS). Although they violate the standard protocol layering, well-configured firewalls proved to be a good protection from the outside intrusions. However, the practice shows that the term "well-configured" is often not given its full dimension — at least until the first hostile attacks.

Two or more localized networks can be connected together by means of a safe channel. We have already stated that a safe communication channel can be established via the unsafe Internet by the use of cryptography, i.e. by the use of safe protocols (2.4). In the general situation of a distributed information system, requiring a complete and integral security, the systems such as Kerberos are to be implemented [15]. Besides the authentication, the appropriate authorization of participants should be performed (see footnotes above for disambiguation and also 3.1). The authorization assures that a participant can access only allowed resources, and execute only allowed actions within the system, in a time-limited schedule.

The firewalls, backed-up with such secure authentication and authorization mechanisms, allow much greater flexibility and connectivity of proprietary networks to Internet, while maintaining high security.

2.7 Summary of the technology related security mechanisms

As a conclusion of this chapter we give the outline of the technology based security mechanisms [16]:

- Firewalls for end-connection to network protection, and intrusion detection systems;
- Proxy servers for access management;
- Content managers for control of the data brought into sent out of the information system;
- Virus protection tools for incoming and outgoing emails and files;
- Service monitors for checking of the service usage, and early detection of the hostile procedures;
- Fail-over systems, to alleviate the loss of availability;
- Encryption implemented in the online applications (EtE), and/or applied to sensitive files;
- Authentication systems: passwords and IDs, physical tokens, cryptographic certificates;
- Digital signatures for verifying the sources of Internet contents.

3 Security and trust

After studying the basic technical aspects of Internet security solutions, we should be in a position of a "well-educated user". Such user can more easily comprehend the security capabilities and remaining risks, and also answer the questions posed in ch. 1.

Now we can turn to the general security aspects which are independent not only of the communication channel in use, but also of the human activity taking place. The security issues of Internet are fundamentally not different from those in the other communication channels. They are just more complex and more important, primarily due to the following facts:

- The use of digitalized data, which can easily be:
  - *modified*, *altered*, *copied*, *replicated*, *distributed*, etc (confer also the discussion in 2.2) as a result of the corresponding malicious activities:
  - *data alteration*, *counterfeit*, *plagiarism*, "*spamming*", etc.
- The use of global, diversified, network:
  - with multi-layered structure that multiplies the points of intrusion, and makes it harder to analyze and control the weakest links;
  - which lacks the global security standardization and implementation;
  - which often lacks the equal legal and ethical support from other communication channels and social institutions.

3.1 The pillars of security

To concretize our discussion, we start by outlining the well known pillars of security. These are:

1. Authenticity — the ability to prove the identity of communicators (confer also the footnotes in 2.3);
2. Secrecy — the ability to keep the information secret from all unwanted parties;

3. Integrity — the ability to keep the information identical to original, i.e. to keep it whole and nothing but the whole;
4. Privacy — the guarantee (a set of rules, policy) that the gathered information will be used confidentially, only by the agreed persons, and only for the agreed purposes;
5. Non-repudiation — legal obligatoriness of performed transactions, and ability to provide undeniable, legally accepted proves of the topics 1 to 4 above.

The above requirements are endangered by the following ***security threats:***

1. Unauthorized acquisition of information, or *loss of secrecy*;
2. Unauthorized modification of information, or *loss of integrity*;
3. Unauthorized decreasing of functionality, or *loss of availability*;
4. Unauthorized loss of control and supervision, or *loss of responsibility* (a situation when everything becomes available, with no limitations and restrictions, and when no one is responsible for the condition of the system).

The threats are to be answered by appropriately implementing the five security pillars, described above, which ensure the following aims:

1. **Secrecy protection**. The message contents must stay secret to everybody but to the trusted partner(s) to whom the message was intended (*contents secrecy*). Also, the transmitting and the receiving side must be able to stay anonymous (*participation secrecy*).
2. **Integrity protection**. Every manipulation with the contents of the message, with the intention to alter and modify it in any way, must be discovered and treated accordingly, in order to reverse the message to its original state, or at least to indicate that it was being corrupted.
3. **Availability protection**. The communication must be available to all the users who demand it, under the condition that their access rights are granted. In other words, they must have proper communication rights according to the system security policy.
4. **Responsibility protection**. This can be further described in the following three points:
   - 4.1 The receiver of a message must have a possibility to prove some third party (e.g. legal authorities) that the defined entity did send her or him the message.
   - 4.2 The transmitter of a message must be able to prove the transmission of the message and the authenticity of its contents, and, if necessary, to further prove that the receiver has received the message.
   - 4.3 Users (customers) cannot deny their obligation to pay for the services, once the provider has sufficient evidences for administering the services in an agreed way.

3.2 Security principles

There are two very general and fundamental security principles to be obeyed:

- **The weakest link principle**: the system is as strong as its most insecure part. A secure system must be nearly equally strong in all of its components, since attacker needs a single (weak) point for breaking in the system.
- **Every user of the system, from both, the outside and the inside world, is a possible attacker**, or can intentionally or unintentionally help some other attacker. A single non-secure point of intrusion, like a personal computer of a negligent user on a local network, seriously weakens the entire system [16].

The second principle could be derived from the first, but is nevertheless stated explicitly to emphasize the human aspect of threats. The problem is to assure validity of the principles in every single component of a complex communication system, such as an online network application.

As an example we may quote the findings that although about 3/4 of attacks come from outside of the firewall, the most damaging and hardest to recover from, are those that come from inside [4, 17], and that can be attributed to the human security aspects.

In chapter 2 we have shown that the technological and technical grounds for implementation of these security aspects to Internet communication and business do exist. But the scientists, engineers and technicians cannot do the complete job even if, by a miracle, they would be able to provide a technically perfectly secure channel. As is already mentioned, the participation and help of all the participants, other professionals, and the whole community is needed. This is yet another illustration of the multidimensionality of the security field.

3.3 Security perception

The above described mechanisms are necessary prerequisites for the development of *trust*. We shall define trust as: *a certainty of some preferred outcome in the future* [18]. While the security can and must be related to the technological and other infrastructure (legal, social), the trust is a notion of humane and psychological nature. It is established on:

- Continuity of regular, desirable behavior of the surrounding;
- Help of the confidential people and institutions;
- Individual knowledge and ability to control the situation.

These three components of trust are overlapping. The continuity of regular behavior depends largely on the functionality of the surroundings. In the technical environment such as Internet, the regular behavior is maintained by technical and organizational procedures. The latter two points are of typical human character, highly dependent on the user's general education as well as on her (or his) knowledge of the Information Technologies and the security issues.

Somewhat peculiar interpretation of trust in the context of security is found in [19]. Trust of a communicator is defined as believing in a positive outcome of a transaction only in the case of lacking certainty. According to this line of reasoning, when the certainty is big there is no need for trust, since one can count on assuredness. The bigger is trust (put in something), the bigger is (potential) risk. This may be a good observation, emphasizing some fine altruistic and benevolent qualities of the term, though, to turn things around, it may be hard to convince someone to put his or her trust in an insecure thing! Our simple model will follow the common notion of the word.[§]

Contrary to trust, distrust is caused by:

- Discontinuity of acceptable behavior;
- Continuity of unacceptable behavior;
- Helplessness.

Trust should rely on the achieved security level. Since it is a highly individual notion formed by rational and irrational human factors, it is the ***perceived security*** that must be considered. If we denote the achieved security level by S, than perceived security S_P should be some function of perception of S:

$$S_P = P(S)\ . \tag{2}$$

The situation is illustrated in Fig. 1, with three different, arbitrarily chosen perception functions. Perception P_1 is the ideal realistic perception for which $S_P = P_1(S) = S$, i.e. P_1 is the identity function. This corresponds to the perception of a knowledgeable and well informed user. Perception P_2 is "consistently" pessimistic, and thus still linear, while P_3 is optimistic perception with nonlinear response.

It is clear that many other variables and parameters, besides the security itself, can and does influence the perceived security. The net contribution of all of them still results in some function similar to those presented here. The goal should be to exclude all irrelevant factors and achieve realistic security perception.-

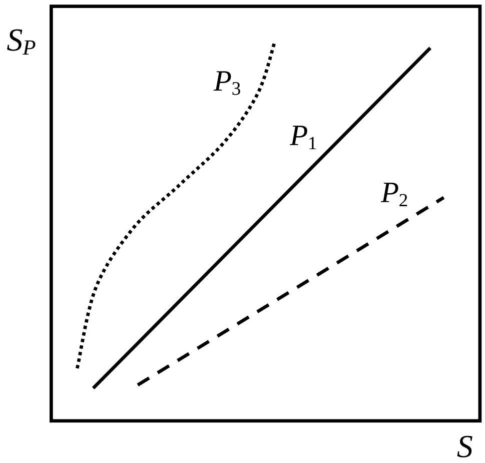

Figure 1. Security level *S* and its perception S_P. Perception P_1 is realistic, P_2 consistently pessimistic, and P_3 optimistic and nonlinear.

3.4 Trust

After our deliberation, the trust T should be some rising function of the perceived security S_P:

$$T = f(S_P)\ . \tag{3}$$

This is illustrated in the Fig. 2. Here an arbitrary rising function is drawn to serve this short discourse. Though quite abstract and without any quantitative ambitions, the graph offers a visualization of the security-trust relationship. In the first approximation, a simplified linear proportion can be considered:

$$T \sim S_P\ . \tag{4}$$

If the function like drawn is assumed, the point A corresponds to a security-unfounded trust (reflecting the previously discussed peculiar interpretation of the term), and C is a point of unnecessary caution. Point B would present some "realistic trust" –- of course, presuming that the function $T = f(S)$ is correct.

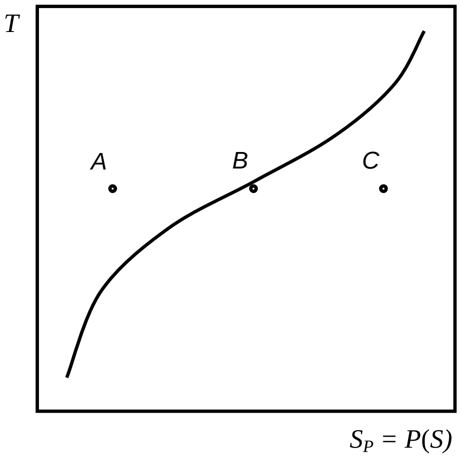

Figure 2. The relation between trust *(T)* and security perception *(S_P)*. The latter should be a realistic estimate (function) of the true security level.

The quantity of trust can be conceptualized through its connection to a certain use or application, and the corresponding risks. For such risks we plan the adequate security. As a general rule, the security costs should be some (considerable) fraction of the risk estimates. Let's say that 1/10 is a good starting point. Specialized IT security companies will suggest more precise investment figures. The costs of all possible damage should be accounted for, including the loss of revenues because of the lost client trust, which should include the costs of rebuilding that trust.

The world outside the communication channel should also be safe, to at least the same extent. In other words the scientific community is obliged to say that:

- The **technical security infrastructure** given solely by proper technology and adequate technical solutions --- is not enough!

The technical security infrastructure should also be complemented by the **human security aspects**:

- Adequate legal support, effective judicial and democratic system;

[§] According to Random House Unabridged Dictionary, trust is: "1. reliance on the integrity, strength, ability, surety, etc., of a person or thing; confidence. 2. confident expectation of something; hope. …

- Community consensus on the security and technology standards, user education, social and ecological issues, etc…

Nothing of the above can be neglected, as it often happens in practice. Give the best available, and the most user-friendly technology to uneducated people, and the problems will arise. Give powerful technology to underdeveloped, or even worse, ill-developed society, and you can be sure that all kinds of communication system abuses will occur because of several possible reasons. This is in accordance to the often quoted fact about the Internet security chain: *humans and human-related aspects are the weakest point* [1].

Regarding the high level of technological development in all spheres of human society today, the above statement is probably true in other communication channels, and also in all other human activities. This dichotomy or, better to say, the dialectics, between the technical and human aspects of communication channels deserves a more detailed deliberation in a separate paper.

4 A glimpse to the present state

Ten, fifteen, years ago, the skeptics would insist that that Internet security infrastructure still requires improvements in the consistent implementation of technical solutions, and much, much better support in the human-related spheres. The use of Internet for delicate communication and pricy transactions was considered too risky and was not recommended. In the meantime, the trends and practice showed them to be wrong---if not in predicting many of the risks and possible problems---than in the tempo at which almost all human activities, from all realms of life, transferred to the ubiquitous use of Internet.

The public trends were positive and enterprising. Even the critical security applications are not exemption from this conclusion, today. ***To not use all the advantages of the Internet, seems like a waste of the great opportunity!*** Such a pro-active public attitude did boost the ICT security sector, because practice called for the immediate implementation of the theoretical solutions, improvements of the global technology standards, and even for the cooperation of the local authorities and institutions in providing better and safer business environment.

4.1 Migration to Internet

The fast development of various kinds of online business communications is for sure witnessed by many of us during the last decade and a half. We got used to the comfort and efficiency of a myriad of Internet services, like: E-banking, E-trading; Direct payments, Internet auctions, B2B communications, etc, even when well aware of the possible risks. These online transactions have broad financial range, and the corresponding broad range of security risks: from a few dozen of EUR or US$, up to hundreds, thousands and much more; from the low level threats of the well-known fraud scenarios, up to the extreme level threats of hacking experts. But we, as “knowledgeable users”, expect that the security solutions are tailored and maintained according to the needs, and that the side assurance mechanisms, starting from the legislation and good practices of the service providers, will protect us from losses. We, as “knowledgeable users”, should also check every now and then the validity of all the security assumptions and expectations.

Today it is more than obvious that all of the above listed Internet business activities are here to stay. They will not decrease in volume, just the opposite. The reasons are obvious:

i. The omnipresence of Internet today is simply dictated by the advantages that it offers, resulting in consumers’ needs and habits in all spheres of life.

ii. For most of the users the losses from frauds are within tolerable limits. Switching to other ways of communication and transactions would cost even more in terms of time and money spent, and again would not guarantee the risk-free operation.

iii. The user experience, practice and reports show that the frauds are not fundamentally Internet-generated, nor solely Internet related, although some of the Internet aspects and features are prone to easy-to-be-performed immoral and illegal acts. But these happen almost proportionally in all other communication channels.

Stated shortly, if ten years ago the question was for which communication and business activities to use the Internet and for which not, nowadays the only question left is how to achieve a sufficient security level for just about every kind of online activity we can, and will do, on Internet.

4.2 The need for relevant statistics

Aside from the fact that online business communication is rolling and cannot be stopped, a serious approach requires in detail statistical analysis as a ground for further discussion and conclusions.

However, such statistics is still missing. Most of the companies, especially those with large transaction volumes, consider these data as highly confidential. The consumer trust could be ruined if the users find out that the security was too low (confer Fig. 2, eqs. 2 to 4). So, even if attacked, the big companies would be solving their problems by themselves, as far as they could. This behavior origins from the early days of the Internet business, when online transactions still had to prove its reliability. As the trust of majority of online customers is already won and their habits generally established, one would expect that more relevant data about the frauds and losses are to be available for broader public.

Some of the companies involved in providing security solutions realize the importance of raising the public awareness by informing them objectively. RSA and CyberSource are good examples [20, 21]. The latter is one of the rare companies providing truly relevant statistics of the lost revenues due to online frauds. Based on this, we have estimated the overall security risk at 0.1% of the total transaction volume, which is close to the order of magnitude of the risk in the offline activities. A more detailed insight and support to this conclusion deserves a separate topic.

According to our investigation, besides the mentioned CyberSource report, not many others, if any at all, are open to public. On the other hand many governmental and nongovernmental organizations, like Fraud Watch and Internet Crime Complaint Center (IC3), do contribute to the public awareness by educational activities and by publishing the fraud statistics reported to them by Internet users.

5 Conclusion

The Internet has grown from an (idealistically) open, free, and insecure place in its beginnings, to a (realistically) less open and free, but potentially much more secure communication channel. We have outlined its existing security technology infrastructure based on the EtE cryptography concept in the Application layer, and discussed the whole palette of security solutions.

These solutions must enable the realization of the security pillars: authenticity, secrecy, integrity, privacy, and non-repudiation. They must prevent, or at least, make futile, the security threats which endanger the mentioned pillars. Upon the well, multilaterally designed, security infrastructure, which, besides the technological solutions includes the important human aspects–-like the instruments of financial and legal protection, the user trust is built. The trust is proportional to the perceived security level. For the latter to be realistic, a proper education, as well as objective, accurate and relevant statistics is needed.

Of the above requirements, the technological solutions are available for quite some time. In practice, however, the problems of consistent implementation, constant maintenance and improvement remain. The human security aspects are also improved, both, in the legislations on the national level and through the security policies of international e-commerce corporations.

To complete this analysis of the Internet security, the relevant statistics should be involved. It must give us better insight of the risks of the particular online activities, as well as to give us an overall risk estimate. A preliminary investigation shows that these risks are similar as in other communication channels.

Furthermore, the Internet and its security aspects can serve us to get a better insight into the problems of general communication channels. Its human versus technological aspects is a topic that deserves further investigation and will be presented in another article.

Acknowledgments

The first author wishes to thank B. Aurer and Z. Hutinski whose ideas and support initiated this paper. We would also like to thank to V. Sac for his valuable comments on the technical aspects of computer networks.

References

1. Müller, G., Rannenberg, K. editors (1); Multilateral Security in Communications, Vol 3, Addison-Wesley, München, 1999.
2. Müller, G., Rannenberg, K. editors (2); Multitlateral Security –Empowering Users, Enabling Applications, The Ladenburger Kolleg "Security in Communication Technology" Annex in [1], 563-570.
3. RFC-1122, Requirements for Internet Hosts -- Communication Layers, R. Braden (ed.), October 1989.
4. Tanenbaum, A.S., Wetherall, D.J., Computer Networks, 5th ed., Pearson, Boston USA, 2011.
5. Simmons, G.J.; Contemporary Cryptology, The Science of Information Integrity, IEEE press, New York, 1992.
6. Netscape, DevEdge Online Documentation, Introduction to Public-Key Cryptography, 1998, http://developer.netscape.com/docs/manuals/security/pkin/index.htm.
7. The SSL Protocol, Version 3.0, http://www.freesoft.org/CIE/Topics/ssl-draft/INDEX.HTM
8. TLS, RFC 5246 http://tools.ietf.org/html/rfc5246
9. The SSH Protocol, http://www.snailbook.com/protocols.html.
10. Unruh, W.; PGP attacks, 1998, http://axion.physics.ubc.ca/pgp-attack.html.
11. Engelfriet, A.; The comp.security.pgp FAQ, 1998, http://www.uk.pgp.net/pgpnet/pgp-faq/
12. Esslinger, B., Fox, D.; Public Key Infrastructures in Banks — Enterprise-wide PKIs, in [1], 283-300.
13. McDermott, J., Attack-Potential-Based Survivability Modeling for High-Consequence Systems, Proc. Third IEEE Int. Information Assurance Workshop, 2005, Washington, DC, USA, J. Cole and S. Wolthusen eds.
14. Riordan, J.; Patterns of Network Intrusion, in [1], 173-186.
15. Kerberos V5 Installation Guide, MIT, 1990 – 1996 http://www.lns.cornell.edu/public/COMP/krb5/install/install_toc.html.
16. Armstrong, I.; Web Commerce – Trading Securely, Security Magazine, October 2000, http://www.scmazine.com/scmagazine/2000_10/feature.html.
17. [x] Verizone Business, 2009 Data Breach Investigations Report, Verizon, 2009.
18. Müller, G., Reichenbach M.; Sicherheitskonzepte für das Internet, 5. Berliner Kolloquium der Gottlieb Daimler – und Karl Benz-Stiftung, Springer-Verlag, Berlin, 2001, section 4.3.
19. Braczyk, H.H. et al.; Trust and Socio-Technical Systems, in [1], 425-238.
20. RSA, The Security Division of EMC^2, www.rsa.com.
21. CyberSource Corporation, www.cybersource.com.