Mobility and Spatio-Temporal Exposure Control

Exposure Control as a Primary Security and Privacy Tool Regarding Mobility, Roaming Privacy and Home Control

Geir M. Køien

University of Agder, Norway; e-mail: geir.koien@uia.no

Received 15 January 2013; Accepted 17 February 2013

Abstract

Modern risk assessment methods cover many issues and encompass both risk analysis and corresponding prevention/mitigation measures. However, there is still room for improvement and one aspect that may benefit from more work is "exposure control". The "exposure" an asset experiences plays an important part in the risks facing the asset. Amongst the aspects that all too regularly get exposed is user identities and user location information, and in a context with mobile subscriber and mobility in the service hosting (VM migration/mobility) the problems associated with lost identity/location privacy becomes urgent. In this paper we look at "exposure control" as a way for analyzing and protecting user identity and user location data.

Keywords: exposure control, vulnerability, risk, identity privacy, location privacy, home control, mobility, cloud, roaming privacy.

1 Introduction

Controlling the degree of "exposure" is one way to reduce risk. If secret and/or sensitive information is exposed then it is more susceptible to being exploited in some way. If we can reduce or eliminate the exposure then the

Journal of Cyber Security and Mobility, Vol. 1, 377–399.

corresponding risk will be reduced or even eliminated (for that particular case).

In this paper we will investigate the concept of spatio-temporal exposure control. Our contexts is users on-the-move (mobile phone/laptop/pad). However, these days it may not only be the user that is on-the-move. Hosted services may also be on-the-move, and cloud services are an example of this. VM migration is already a well-established concept and VM mobilities have also been proposed and discussed in the literature.

Some services may then even move along with the user. In the future one may even subscribe to "follow-me" services, though the most likely seems to be that "follow-me" would be a quality-of-service attribute. Services that need low-latency and services that needs to stay within the same jurisdiction as the user may benefit from a "follow-me" feature.

Mobility is the order of the day, and we should expect this to affect the intruder(s) too. In a geographically distributed environment it may be necessary for the intruder to move alongside its targets, or otherwise it may fail to intercept communications, etc. In this respect it is important for the intruder to be able to distinguish users and services uniquely, and so it will be a goal for the intruder to obtain tracking references to the various objects and entities.

1.1 Exposure Control

Risk analysis methods and the corresponding countermeasures and mitigation is an important part of systems design, configuration and deployment. Modern methods like the TVRA methodology (see Section 2) represent a fairly complete approach to risk assessment, but there is still room for improvements.

The "exposure" an asset experiences plays an important part in the risks facing the asset. The exposure is, technically speaking, not a risk, but it certainly can put vulnerabilities and weaknesses into focus. Thus, increased exposure will increase the probability that vulnerabilities and weaknesses are uncovered. In this context we propose that exposure control mechanisms will be a useful tool in controlling the risk.

1.2 Home Control

The "Home Control" concept originates with cellular operator community and has had a particular standing within the North American operator com-

munity [5]. The basic problem that faced the cellular operators was that the classical roaming model was, with regard to trust, a very naive model.

The cellular roaming model is a model with extensive delegation of responsibilities to the visited network. The delegation even extends to the authentication and key agreement protocol; the sessions security credentials are simply forwarded to the visited network. Even worse, the forwarding of the security credentials is potentially not bounded to any authentication event. That is, the visited network may receive the credentials at time `T1` and only use the credentials at time `T2`. The home network is normally not alerted to the authentication event at time `T2`, and thus the home network is functionally offline with respect to authentication of the subscriber [4].

Needless to say, the cellular model leaves a lot to be desired with respect to home control; the home network, and for that matter the subscriber, must trust the visited network to an unreasonable degree with respect to incurred charges from service consumption, etc. The home network has almost no way of verifying that the subscriber has consumed the services since it very seldom is in direct contact with the subscriber when the subscriber is roaming.

This is unsatisfactorily seen from the home network perspective and there is a clear need for the home network to have more control over the authentication. Tighter control over the associated charging is in place, one option is to require near real-time exchange of charging data, but this is still a reactive fact measure.

Home control classification:

- *Pro-active Home Control*
  Deployment of strong 3-way online authentication is a pro-active security mechanism. Access control functionality is another example. Other schemes that aim at prevent problems from ever occurring would also be classified as a pro-active mechanism.
- *Re-active Home Control*
  Real-time charging and anomaly detection schemes is a re-active security mechanisms. Basically re-active mechanisms must have a strong and focused detection capability in addition to an ability to react adequately to the detected incident.

Both pro-active and re-active schemes will have their place in a security architecture.

1.3 Home Control for Cloud Services

The home control concept found in the cellular roaming context has also relevance for cloud services. For instance, in a public cloud environment the cloud service operator has a similar role to the visited network in a cellular environment. The subscriber and the home network will be similar to the VM initiating user and the organization he/she is associated with (and which has the agreement with the cloud service operator). The mapping is the following:

- Cellular Subscriber $\cong$ VM initiating user (`USR`)
- Home Network $\cong$ Service Subscriber Entity (`SSE`)
- Visited network $\cong$ Cloud service operator (`CSO`)

The Service Subscriber Entity may be identical to the VM initiating user, but it may for instance also be the employer of the user.

The problems with lacking home control is also quite similar, and we note that in a basic configuration the `USR/SSE` has very little control over the submitted VM, the associated data and the outcome of the VM execution. To some degree the problem can be solved the same way as one did for the cellular roaming case, namely to more or less blindly trust the cloud service operator to protect the program/data and to carry out the requested operations as intended. However, it should be clear that while the naive trust model underlying cellular roaming have worked well it is quite inadequate for many, if not most, scenarios which involves processing of confidential and/or otherwise sensitive data.

1.4 Spatio-Temporal Contexts and Mobility Model

The overall context described and discussed in this for mobile/cellular subscriber and for mobile hosted services. The mobile hosted services are VM based services where one may expect VM migration or even VM mobility. The intruder may also be mobile, or even geographically distributed. Figure 1 depicts a possible mobility model. With respect to temporal issues we expect all contexts to be temporally contained, but also that context renewal is possible.

1.5 User Privacy and Identity Protection

Thus, for both cases we have that the location is variable parameter. Another part of the context for our investigation is user privacy. Given that we deal with mobility it is no surprise that location privacy is of interest, and associ-

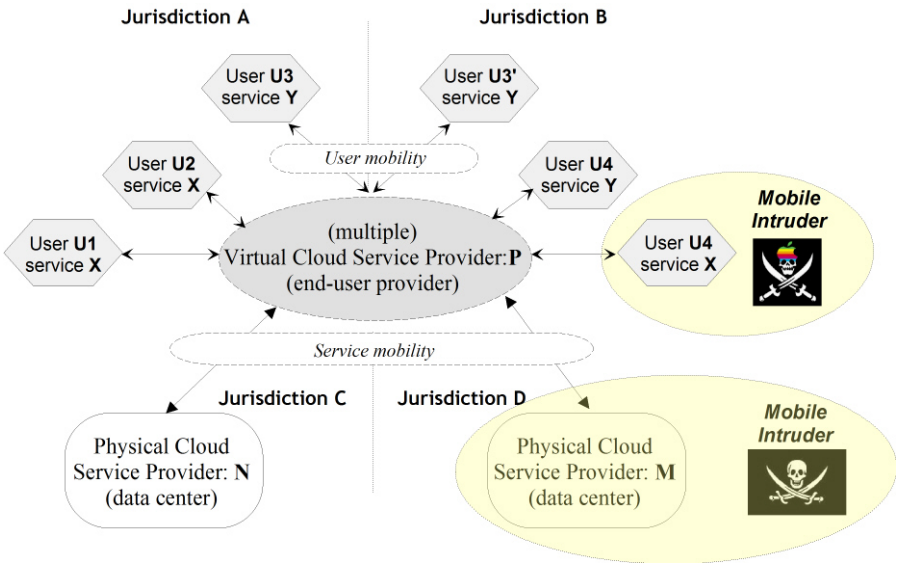

Figure 1 Generic 3-way mobility model.

ated with it we have identity privacy. Data privacy, transaction privacy, etc., also comes into play, but in this paper we primarily deal with identity and location privacy. Privacy, of course, may very well be an end to itself, but we note that lack of credible privacy may easily lead to other security problems.

Identity theft is a growing concern and while it may have received more attention a few years ago, it really does have an impact. According to *The New York Times* (2011/02/09, "The Rising Cost of Identity Theft for Consumers" [19]) the reported number of incidents in the U.S. actually fell in 2010 by approx. 28%, but the cost associated with identity theft still rose. A staggering 8.1 million adults in the U.S. were victims to identity theft and the associated cost has been estimated to be approximately $631 on average. This number did rise sharply from 2009 when the cost was only $387 on average, and the total is now in excess of $5 billion. Similar numbers have been reported elsewhere and in the U.K. the reported numbers were an accumulated cost in excess of £2.7 billion and it affected more than 1.8 million people [20].

Measures that reduce the risk of identity theft therefore clearly seem worthwhile and to limit the exposure seems indeed to be a useful approach.

1.6 Identity Theft

Identity theft is not specifically considered in this paper. Suffice it to say that if one looks at the impersonation aspects of identity theft, it should be clear that a successful identity theft scam relies upon two factors:

- Knowing the identity/identifier of an entity,
- Convincingly claiming to be the said entity.

To prevent impersonation/masquerade one must then prevent the intruder from learning the identity/identifier and/or preventing the intruder from being able to corroborate the identity/identifier.

From a security perspective alone it is not important to conceal the identifiers, in fact identifiers are almost always presented in plain text in authentication protocols presented in the literature [33]. Strong authentication will effectively prevent masquerade. By strong authentication we must here require that a security context is set up by the authentication procedure and that key material associated with the context is used thereafter to cryptographically protect all transactions between the parties.

From a privacy point of view one should obviously not leak privacy sensitive information like an identity. Short-lived transient identities may not matter that much, but permanent or long-lived identifers may allow an intruder to track the target entity. Of course, to claim that exposure of short-lived identifiers does not matter requires qualification. What is short-lived supposed to mean? Furthermore, we *must* require that there is no apparent correlation between the various identifiers used by the same entity. To have a string of short-lived but obviously connected identifiers will not do, as the emergent property would be that of a long-lived identifier.

We should also mention that one may benefit security-wise too from not exposing the identifiers unduely. This is mostly due to imperfect security mechanism, implementation weaknesses and system architecture constraints that sometimes will allow an intruder to potentially gain a weak advantage if he/she knows a subscriber identity. We therefore claim that identity exposure control will also have tangible security benefits as a defense-in-depth type of protection scheme.

2 Brief Introduction to Threat Vulnerability and Risk Analysis

2.1 Next Generation Network

In order to put "exposure control" in context we shall briefly investigate the Threat Vulnerability and Risk Analysis (TVRA) [1] concept. The TVRA methodology was developed by the ETSI TISPAN project for the so-called Next Generation Network (NGN) architecture. ITU-T defines NGN to be:

> A Next Generation Network (NGN) is a packet-based network able to provide services including Telecommunication Services and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent from underlying transport-related technologies. It offers unrestricted access by users to different service providers. It supports generalized mobility which will allow consistent and ubiquitous provision of services to users.
>
> `www.itu.int/ITU-T/studygroups/com13/ngn2004/working_definition.html`

The ITU-T defined NGN systems architecture has and will have a huge influence on the major core networks and the main access networks.

2.2 Critical Infrastructure Protection

The NGN concept must also be seen in a societal context with increasing dependency on information and communications technology (ICT). In this context the resilience and dependability of the NGN infrastructure becomes crucial, so much so that one has defined the concept "critical infrastructure (CI)". A number of papers and reports has been written about critical infrastructure protection (CIP) and Elsevier has even launched a scientific journal catering to this topic (*International Journal of Critical Infrastructure Protection* (IJCIP)). This paper is not about CIP per se, but we note that exposure control can easily fit into the overall CIP concept.

2.3 Vulnerabilities, Threats, Risks and Threat Agents

Given that an NGN infrastructure is a critical component in a modern society it is necessary to ensure that it is dependable and secure. Within the ETSI TISPAN project one has developed a new methodology for analyzing

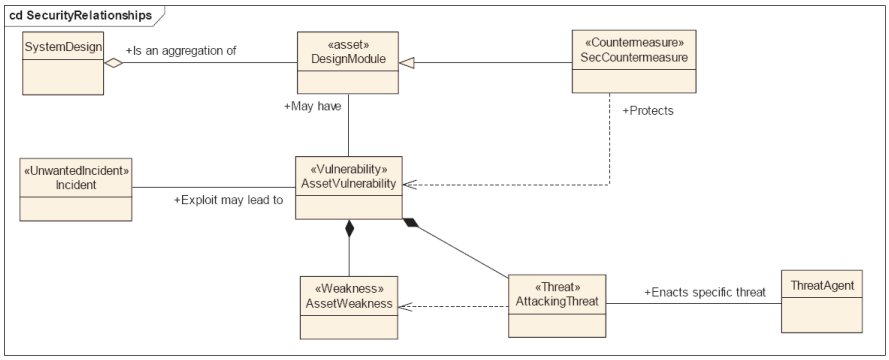

Figure 2 Generic TVRA model.

vulnerabilities, threats and risks associated with NGN type of networks. The TISPAN initiative also encompasses countermeasures and cost-benefit analysis, etc., for the various cases [2]. We shall now briefly outline the TVRA concept [1].

2.3.1 Generic TVRA Model

We have that an *asset* is an object of value that needs to be protected. In a system there may be unwanted/undesireable events concerning an asset. These events are denoted as *incidents*. An incident occurs when a `vulnerability` is exploited. A vulnerability is then a *weakness* which may be *attacked*.

The weakness/vulnerability may exist without there being any incidents, but given a knowledgable *threat agent* the weakness/vilnerability may be exploit and used in an attack. Figure 2, transposed from [1, fig. 4], depicts the Generic TVRA model.

2.3.2 Security Objectives and Threats

In the TVRA model on defines four primary security threats and five primary security objectives. Primary threats:

- Interception,
- Manipulation,
- Denial of Service (DoS),
- Repudiation (sending and/or receiving).

The security objectives do not correspond directly with the threats, but there are obvious relationships. The primary security objectives (known as CIAAA):

- Confidentiality,
- Integrity,
- Availability,
- Authentication,
- Accountability.

2.3.3 Weaknesses, Vulnerabilities and Threats

A precondition for a threat, in the above model, is to have a threat agent. For a large system it is naive not to assume that the threat agent, also known as intruder, adversary, enemy or even hacker, is present. There exist many different types of threat agents, ranging from spectacularly powerful intruders to opportunistic and less resourceful legitimate users that simply try to elevate their access rights beyond what has been agreed. One examples of these is the classical Dolev–Yao Intruder [12] and in [13] one defines a set of intruders based on their capabilities and financial strength. In [14] one further discusses the computational strength of attackers in context with Moore's "law".

So, we assume threat agents to be present. Some of these agents will be powerful and some will be less so, but the lesser agents may be numerous and may in the end prove to be a larger problem for the overall system.

A real-world system will have weakness and vulnerabilities. Some of these weaknesses and vulnerabilities are simply due to weak design or erroneous implementation, while other arise due to inescapable complexities or due to design decision that give priority to certain features over other features. According to Anderson's classical "Why Cryptosystems Fail" [15] one should also assume that quite a few of these weaknesses are due to misunderstood security objectives, to inadequate threat models and to misguided trust assumptions.

Whatever the reason or cause, the vulnerabilities and weaknesses exist in the system and they will be susceptible to exploitation by an threat agent provided that the vulnerabilities/weaknesses are visible to the threat agent. In this context we argue the case that "exposure" should be included as a class of vulnerability and that "exposure control" should be an independent counter-measurement in an extended TVRA method.

2.3.4 Conflicting Incentives

Why do we carry out risk analysis activities? Obviously, to identify risk and to reduce and mitigate it as we see fit. However, what is a risk or liability to one party is an opportunity to another party. In terms of privacy, it should be clear that private information has value to more than one party. Unfortunately, the

there will often be clear conflict of interest, and this is dependent for privacy sensitive information.

An example would be web surfing and searching. You may want to remain anonymous while Google, Facebook, Instagram and other services will potentially stand to make profit from knowledge about you and your habits. So there may very well be conflicting incentives during the web transactions. The Conflicting Incentives Risk Analysis (CIRA) methodology is one way to capture this [34].

A risk analysis methodology should be able to capture and cater for conflicting incentives and interests to be truly useful. The TVRA methodology does not currently cover this, but it should be possible to extend it to cater for those needs too, perhaps by including CIRA methods.

3 Mobility and Migration

It goes without saying that cellular subscribers can experience full mobility. Seamless mobility is also a standard service in cellular systems. Traditionally the functionality has been limited to mobile phone handsets, but nowadays mobile termination (MT) units are commonly integrated into laptops, tablets and other gadget. The mobile device may of course also be an embedded device, and we may therefore potentially include all Internet-of-Thing (IoT) devices. The distinction between traditional cellular services and other wireless service are also blurred and more or less meaningless to the customers. Thus, we can safely postulate that users with laptop/smartphone/ipad and other gadgets will, as the default assumption, be mobile subscribers in the sense that they can obtain IP connectivity and that they routinely are on-the-move while being connected.

With the inclusion of IoT devices in the equation we must cover several communications scenarios:

- *Human-to-Human* We note that while the communications may logically be human-to-human it may certainly be conducted and facilitated by mobile devices at the lower layers.
- *Machine-to-Machine* Embedded devices are quite often wirelessly connected. It those cases one must assume mobility to be the norm. As stationary wireless device can safely be modelled as a mobile device with a special case of zero velocity.
- *Human-to-Machine*

The interface must be adapted to humans, but apart from that human-to-device communications need not be special at all. Again, we shall assume wireless communications to be the model.

Technically speaking there isn't a big difference between the mobility handling in the above scenarios. There may be humans involved on causing the mobility, but the technical realization of mobility handling at the lower layers will invariably be handled by some mobility management machinery.

In the reminder of this section we investigate service mobility in the guise as VM migration/mobility.

3.1 Physical VM Migration

Live migration is by now a standard option in most cloud services. Basically it allows a server administrator to move a running VM or application between different physical machines while providing uninterrupted service. Live migration requires that allocated memory, storage, and network connectivity of the VM is successfully migrated to the destination machine. Seamless migration is defined to be a migration event that is transparent to the services consumer.

Migration is normally considered to be a "local" event in the sense that one normally assumes that both source and target machine is physically close to each other, i.e. within the same data center. That is, one can safely assume that normal VM migration is restricted geographically and generally within the same physical premises. So one does not need to worry about switching country or switching host operator, etc.

3.2 Physical VM Mobility and VM Roaming

VM mobility is somewhat more of a novelty, but it is not a new concept [23]. In VM mobility the VM is moved beyond the traditional "local" boundaries and the mobility is not per se limited in physical distance. In practice one cannot have full service continuity for prolonged relocation procedures, but this would of course depend crucially on required service response times and on the quality of the connection (bandwidth and latency).

The case argued in [23] is for very low latency services and where the executing VM needs to be in the physical vicinity of the user in order to minimize network propagation delays. Whatever the motivation, it should be clear that techniques that allow VM migration would also allow VM mobility. The upshot of VM mobility is that one cannot be entirely sure that the VM

stays within the same data center and therefore it may potentially move across borders and potentially onto a different hosting environment/service.

We shall denote VM mobility onto a different host environment as *VM roaming*. This will include VM mobility from host `A` to host `B`, where host `A` and `B` is the same company, but located in different jurisdictions.

3.3 Physical Mobility of the Server

The service platform may itself be a physically mobile platform. Laptops, mobile phones and other gadgets may be used as a host platform. These platforms, while somewhat computationally restricted, are plentiful and are themselves mobile. Social networks or corporate networks may utilize the user client platform to host simple cloud services. These services may be private cloud services and they may be specific to a service, but ultimately they could be realized as publicly available generic hosting services.

3.4 Technology Mobility

As of today the VM technologies are fairly specific both to physical host platform and to hypervisor/VM manager type. However, it is of course possible to fully emulate one environment within another, and so a type-X VM can be run on a type-Y environment provided that a X-to-Y emulation layer exists. It is therefore, in principle, possible to have *VM roaming* cases where the VM moves onto a different platform from where it originated. The technology has not reached that level of maturity yet, but if there are strong enough incentives then surely new technology will be developed to allow this to happen.

4 Exposure Control

4.1 The Case for Exposure Control

The concept of exposure control is not directly linked to weaknesses or vulnerabilities, but obviously the less exposed a weakness or vulnerability is the less likely it is that it can be converted into an attack.

Thus, as a means of "defense in depth" [16] exposure control is about reducing the exposure of assets to a minimum. Defense in depth has not the best reputation in academic papers, but some recent papers analyzing threats and attacks have found that "defense in depth" and "security by obscurity" does have merit in the sense that broad sweeping attacks can be prevented and/or mitigated by these tactics [17, 18]. The reason is found in the cost

associated with attacking large populations, but we should warn that these tactics are less likely to be effective against targeted attacks.

Still, we argue, protection schemes that simply aim at concealing the presence or obfuscating the presentee of an entity may be effective against the opportunistic attacks. Since, according to Florêncio and Herley [17] and Pavlovic [18], there is reason to believe that these attack are the most common ones, it makes a lot of sense to employ defense tactics that limit and control the exposure of assets.

4.2 Cryptographical Exposure Control

Exposure control is not a new concept per se and in cryptography and information security there are well developed notion of exposure with respect to cryptographic keys and to the amount of ciphertext that should be encrypted under the same key.

A secret cryptographic key has only a limited amount of entropy and while one relies on effective cryptographic primitives to mask any correlation between the key and plain text data, there will inevitably be information leakage. This leakage cannot be avoided and one must therefore restrict the use of a key so that it does not get too exposed. Exposure and information leakage also happens at the key distribution phase and during storage. The key distribution and key agreement problem can largely be contained with good cryptographic protocols, but the storage problem is harder to solve. A weakness in the hardware platform, any weakness in the system software, the security software or even the application software may leak information about the key. This type of leakage may be entirely independent of the actual usage pattern for the key.

To address these issues cryptographic systems and protocols typically limit the lifetime of secret keys. An example is the IPsec protocol suite where one can limit the "lifetime" for a security association both in terms of usage (no. of bytes/packets) and in terms of passed time (seconds) [21].

In [22] the case is argued for spatio-temporal exposure control and this paper is an important background paper for our investigations.

4.3 Privacy Assets

The primary privacy assets will be permanent identities, the associated location data and of course "data privacy".

Data privacy, in a communications setting, will normally only cover the protocol payload, but what is considered payload is a matter of perspective. At layer **N** the whole of layer **N** + 1 is payload. So, we will refrain from a strict definition, and rather allow "data privacy" to include "all relevant data". With this in mind it is clear that data privacy protection must be implemented sufficiently low in the stack to be able to protect "all relevant data". This is why it is necessary to deploy data confidentiality protection at the link layer in cellular/wireless systems. However, link layer protection is, by definition, limited to the range of the link. Once inside the core network the norm is to protect (aggregated) traffic at the network layer, but this isn't necessarily sufficient as some services are more sensitive than others. That is also why one may need additional protection at higher layers to cover end-to-end aspects.

One may classify the identifiers according to geographic scope and lifetime. An example is shown in Figure 3. A permanent identity will, by definition, by comparatively long lived. It may not necessarily be a secret per se, but is has the potential to be highly privacy sensitive.

Under many circumstances one does not actually use a primary identity for transactions purposes. One may instead use secondary identifiers (numbers, references, addresses) which may be derived from the primary address. Additionally, there may be "emergent" identifiers that may or may not be recognized as identifiers per se, but that may nevertheless be used for tracking purposes by external entities (including our intruder/adversary). Many of these secondary identifiers will be public, but they may also be private. Furthermore, a secondary identifier may have limited lifetime. This may arise out of the given context or it may be by explicit design.

One may also find that there are identifiers that does indicate a class of objects or entities rather than a specify object/entity. However, prolonged use of a class identifier by any specific object/entity may allow for additional data to be associated with the specific instance and so one may in the end derive a unique identity from the context. This "derived identity" may or may not be recognized by the object/entity that it refers to.

Tertiary (transient) identifiers will also exists. These will be short-lived and/or be temporally and/or spatially contained. A typical example would be the `M-TMSI` temporary identity used in LTE networks [6]. The `M-TMSI` is unique within the respective MME area, but needs to be qualified for external use (forming a `GUTI` identifier).

An identity, even a class identifier or a secondary/tertiary identifier, is obviously an assets in our case. The location of the an identified entity/object may also be viewed as an asset. The more precise the location the more valu-

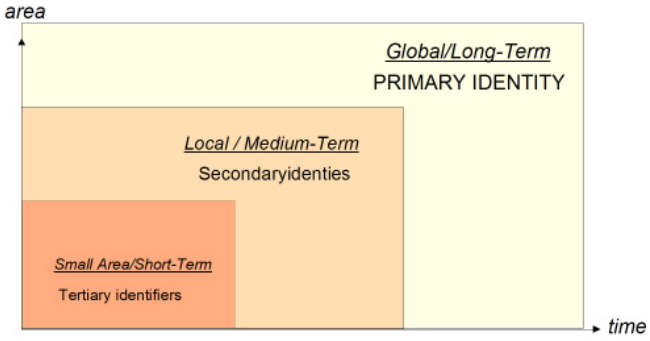

Figure 3 Longevity and scope of identifiers.

able the asset becomes. For mobile subjects to accumulate a time-series of identity/location information is another type of privacy asset, and it is in some cases a more valuable asset. The time-series amounts to tracking information and given tracking information and some traffic data one may easily also arrive at transaction information. Should one be able to also gather user data and associate it with tracking information then one may compose a fairly complete tracking record and this could very well be used for identity theft attacks or similar. As was discussed in Section 1.5 identity theft is a very real threat and the effects of identity theft is amongst the worst both economically and emotionally.

Other identifiers, information and information patterns may also be used as an associated identifier for the subscriber. These may also be abused in identity theft scams. There are many auxiliary identifiers used in a 3GPP system context and they include amongst others the ICCID (smart card ID), the MSISDN number (the phone number) and the IMEI (mobile device serial number), not to mention other non-system identifiers that may be associated with the user/subscriber like various account identifiers (Android/Google, Skype, Facebook, etc.). We shall in this paper limit ourselves to link layer identifiers and then primarily to those associated with setting up an initial security context. The impact of of identity theft, as major source of a privacy intrusion, is discussed in more detail in [24].

5 Privacy-based Spatio-Temporal Exposure Control

5.1 What to Protect

The question is of course not only what to protect, but also from whom one needs to protect the information. When it comes to identity information and to location information we may start off with declaring that "external" non-authorized parties shall not learn neither identity nor location for a user or indeed for an associated (VM-hosted) service. But there are potentially many "internal" parties and not all of them really need to know the privacy assets. So, one must really start off with defining who should have access to the privacy assets. Here we strongly advocate prudence and we advocate that privacy respecting business principles to be used, along the lines of the Privacy-by-Design initiative [25].

5.2 Where Are the Identifers Located

The various identifiers are potentially stored on a lot of different nodes. In a distributed system this means that sensitive information will be stored on nodes in different areas and oftentimes under different jurisdiction. This will make privacy protection complicated, but with respect to enforceability and with respect to trust. To illustrate the complexity we shall briefly outline where some of the 3GPP identifiers are stored. The overview in Figure 4 is by no means intended to be exhaustive. For instance, we have excluded the user equipment (UE) entirely. It must also be mentioned that since the figure contains a mix of 2G (GSM/GPRS), 3G (UMTS) and 4G (LTE/LTE-Advanced) identifiers, credentials and nodes, it is bound to be somewhat imprecise. Inter-generation support (backwards compatibility) for roaming, etc., complicates this picture further.

It should be mentioned too that paging and system access in the 3GPP systems necessarily involves exposure of identifiers. For paging, the identifier will be visible on the over-the-air interface within the whole of the location area (routing area). It should therefore be immediately clear that paging identifiers are widely exposed and that consequently one should *never* page a subscriber with the permanent identity (IMSI). Corollary, there should be a limit to the number of paging events per temporary identifiers too. It should also be clear that quite a few nodes will see privacy sensitive data and that all those nodes must therefore be able to protect the privacy assets.

As Figure 4 shows, the 3GPP systems, with the distributed authentication model, are also vulnerable in that the authentication material and key material

	Serving network					**Home**
	Cell BTS/NB/eNB	**BSC**	**RNC**	**MME**	**VLR/SGSN**	**HSS** (HLR/AuC)
IMSI	Yes	Yes	Yes	Yes	Yes	YES
TMSI	Yes: BTS, NB	Yes	Yes	Not applicable	YES	No
GUTI	Yes: eNB	Not applicable	Not applicable	YES	Not directly applicable	No
MSISDN	Yes	Yes	Yes	Not applicable (except non-ISUP use)	Yes	YES
IMEI	Yes	Yes	Yes	Yes	Yes	Yes
Auth.Info	Challenge-Response data	Challenge-Response data	Challenge-Response data	Full EPS-AV and Legacy contexts	Full triplet and full UMTS AV	YES, ALL
Key material	Yes: BTS,eNB	Yes: GSM	Yes: UMTS	Yes: LTE	Yes: GSM/GPRS and UMTS	YES, but not *AS Security context* in LTE

Figure 4 Distributed identifiers and credentials.

(triplet, UMTS AV, EPS AV) are distributed to the serving network. Exposure of these credentials would make identity theft and impersonation very easy to carry out.

What is not shown in the figure is that the identifers and credential must also be transported between the nodes. So, one also need the communications security to be reliable, available and actually used. In this respect it is clear that the 3GPP security architecture (2G in TS 43.020 [10], 3G in TS 33.102 [10] and 4G in TS 33.401 [10]) is not very strong on requirements on the deployment of the so-called Network Domain Security (NDS/IP) protection (TS 33.201 [9]). There is therefore a high probability that the identifiers and credentials are not protected while in transit. We add here that this may even be the case for information that passes through intermediate networks.

5.3 How to Protect Privacy Assets

Needless to say, this question cannot be fully answered without taking the context into consideration. One needs to define the privacy assets that need protection and one needs to define what it must be protected against and possible also how it is to be used.

For cellular systems one has the potentially conflicting requirement that the home network (HPLMN) needs to have home control while the subscriber (represented by the user equipment UE) needs to have credible privacy. The system access protocols in 3GPP-based system (GSM/GPRS, UMTS, LTE/LTE-Advanced), which includes identity presentation and authentica-

tion and key agrement, typically expose both permanent identities (IMSI) and secondary/tertiary identities (TMSI/P-TMSI/GUTI). See 3GPP TS 33.401 [10] for details on the security part of the access procedure.

One also has the serving/visited network (VPLMN), which needs assurance that incurred costs for service provisioning will be accounted for. In [28, 27] this conundrum is discussed and a solution is provided that does indeed provide credible user privacy and a fair amount of home control. Here the basic idea is that mobile device presents itself with an anonymous pseudo-random subscriber identifier, *ASID*, and an encrypted block *A* that contains, amongst others, the long-term identity. Block *A* is encrypted with the public-key belonging to the home operator (HPLMN), and through the AKA procedure the HPLMN will get assurance about its subscriber while the long-term identity is concealed from both the VPLMN and external parties (the intruder). The VPLMN will get confirmation from the HPLMN that the $ASID$ is representing a recognized subscriber and that the HPLMN will accept charging on behalf of the subscriber.

For *location privacy* vs. *home control* one may additionally use Secure Multi-party Computation (SMC) methods to let the home network question the serving network/user about the location while essentially only providing assurance about location without actually revealing the location. In [28] a demonstration of this scheme is demonstrated through a protocol which solves the so-called point-inclusion problem. The protocol in [28] is reasonably efficient (for an SMC protocol), but it is not too practical and it can be circumvented by a dishonest party. Other solutions exists too, and in [3] several of those are discussed.

The cellular system setting can therefore be said to have some solutions and the solutions are even quite good. Other settings which have solutions include IoT-based cases in which a user may access an IoT-based service without revealing too much private information to the IoT device. The actual requirements will dictate the how one solves the problem; Køien [26] provides one example. Another example is found in [31] where one investigates problems associated with privacy and intrusion detection on a mobile broadband platform.

Cloud service privacy is an area where, to the best of the author's knowledge, there is no truly credible and practical solution available yet. The problem is hard in the sense that a VM executing on a remote platform cannot easily verify it own location. Home control for the VM owner is therefore hard to come by. We may attempt to briefly sketch a way forward here, and it seems reasonable to start off with a requirement for *verification*. In

cloud parlance this is often described as the *remote attestation problem*. That is, there must be some way for the software-only VM to *verify* its identity (ownership) and its location. The location aspect we may be most interested in here may be the jurisdictional location, but whatever way we choose to classify the location information we still need to have means for verification of (hypervisor) claims. Use of Trusted Computing Module (TCM) functionally [30] may be a way forward, but there nevertheless seems to be a need for at least semi-trust in the cloud service provider. On the subject of trust and cloud services there is actually ways for increasing the trust one may have in cloud services [29]. This does not replace the need for "hard" assurance, but may be a useful addition and a acceptable "defence in depth" addition.

If verification is possible then the next logical step for the VM is to apply that information. Specifically, the VM should now attempt to address and comply with the home control policy. Thus, the VM must somehow be able to *enforce* the home control policy. For instance, if the VM detects that it is executing in a foreign (hostile) legislation it may need to shut down or it may need to set up additional security measures, etc. We believe that if verification is possible then enforcement should be possible too. Another complicating aspect here is that the remote attestation must be conducted for mobility cases. If the VM migrates or is otherwise relocated then clearly one must re-attest the platform. If fact, the re-attestation should be performed before the relocation takes place and it would seem reasonable to assume that the current VM host is responsible for verifying the target VM host before actually moving the VM to the target host. Needless to say, mobility/migration should be subject to policy control and maybe even to some emasure of VM home control.

When it comes to generic identity protection we should of course not forget to mention identity management solutions. Many proposals and initiatives exists and there are also several (national) standards available. One prominent initiative is "The National Strategy for Trusted Identities in Cyberspace (NSTIC)" [32]. For many cases the use of identity management schemes is the only way forward and thus the way privacy and security is handled is of the utmost importance.

5.4 Privacy Policy Control

We believe it is essential to provide some means of privacy-based policy control, which may be used for migrating VMs and roaming subscribers. In the 3GPP framework one already has a general scheme called the "Policy

and charging control architecture" for policy control for roaming subscribers (TS 23.203 [7]). The policy control architecture is mostly concerned with charging and QoS aspects. It would be very useful to apply this framework to instruct the roaming partner on how to behave with respect to subscriber privacy.

Exactly how one should best do this remains for further study, but it is immediately clear that policies regarding the frequency of re-assignment of temporary identifiers and of use of NDS/IP protection would be welcome from a security and privacy perspective.

With respect to cloud computing and VM mobility it is clear that it too should be subject to policy control. The VM "handover transfer" should be protected, the target VM should be verified (remote attestation) and the target VM host should be an allowed host. Since tracking of VMs could be an issue it is also important the VM references are anonymity or fully confidentiality protected.

6 Summary

Exposure control is an important aspect of a security architecture. It relates directly to the assets in the system and will as such be part of a generic risk analysis. There will also be exposure control mechanisms, and for privacy these will be associated with various confidentiality services.

Exposure control still needs to be investigated as a part of an extended TVRA methodology. Exposure itself is an aspect of vulnerability and exposure control is a way of mitigating a possible threat. We also advocate to take conflicting interested into account to more accurately reflect the real world. More in-depth work in this direction is recommended.

Exposure control is relative with respect to the asset one is concern about. The asset have different value to the different entities in the system, and there may be conflicting interest here.

In this paper we have primarily looked at knowledge of user identity and user location as the primary assets. Thus, we have in effect investigated privacy exposure control. Privacy is a means to itself and privacy is also a growing concern. Thus, efforts in mitigating privacy problems is clearly is also a means to itself. However, in the literature we saw that privacy intrusions rarely appear to be only an end to itself. In fact, identity theft seems often to be the attacking purpose. That is, identify theft is again the starting points for fraud at large. This is not only a theoretical concern and identity theft related crime has accumulated costs in multi billion dollar region.

For cellular subscriber and for users that needed IoT-based service access we found that solutions existed that would permit credible protection of identity and location. We also saw that identity management solutions exists, and when applied correctly these may indeed also provide a measure of privacy. We note that the user/subscriber does have its own hardware (the mobile phone/device, trusted smartcard (SIM,UICC/USIM)) and that this obviously helps the assurance. Not that one should be too naive here as malware may corrupt the platform, but it obviously is possible to have higher assurance levels when one has control over the hardware platform.

To some extent all the solutions require a level of trust in the participating parties (not always all of them), and this points to the fact the trust management (and associated enforcement mechanisms) is also needed alongside with identity management.

For cases where the mobility is in the hosted service, e.g. in VM provided services, the case is more worrying. Data confidentiality, VM referential identity and VM location may all be privacy sensitive and the VM owner will need a level of "home control" over the VM. Exposure control in this setting is difficult since the VM is all software based and the VM is hosted on hardware which is not under control by the VM or the VM issuer. Thus, there appears that there is no viable way to establish the current status (verification) or to enforce a particular privacy and security policy. That is, use of trusted hardware at the cloud service provider may allow some home control. It should be possible to have remote attestation (verification) and it may be possible to even have a certain level of enforcement.

We believe further research is necessary to conclude on this and we believe that remaining exposure issues can at least partially be solved or mitigated by providing credible trust management solution in conjunction with identity management solutions that emphasizes privacy. Since we are dealing fraud and crime with multi billion dollar interest it seems that contractual matters and legislation that favors safe business conduct also needs to be in place.

References

[1] ETSI, TS 102 165-1. Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Methods and protocols; Part 1: Method and proforma for Threat, Risk, Vulnerability Analysis. March 2011.

[2] ETSI, TS 102 165-2. Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); Methods and protocols; Part 2: Protocol Framework Definition; Security Counter Measures. February 2007.
[3] G. M. Køien. Entity Authentication and Personal Privacy in Future Cellular Systems. River Publisher, Aalborg, Denmark, 2009.
[4] G. M. Køien. An introduction to access security in UMTS. IEEE Wireless Communications, 11(1): 8–18, February 2004.
[5] G. Rose and G. M. Køien. Access security in CDMA2000, including a comparison with UMTS access security. IEEE Wireless Communications, 11(1): 19–25, February 2004.
[6] 3GPP. TS 23.003 Technical Specification Group Core Network and Terminals; Numbering, addressing and identification. 3GPP, December 2012.
[7] 3GPP. TS 23.203 Policy and charging control architecture. 3GPP, December 2012.
[8] 3GPP. TS 33.102 3G security; Security architecture. 3GPP, December 2012.
[9] 3GPP. TS 33.210 3G security; Network Domain Security (NDS); IP network layer security. 3GPP, December 2012.
[10] 3GPP. TS 33.401 3GPP System Architecture Evolution (SAE); Security architecture. 3GPP, December 2012.
[11] 3GPP. TS 43.020 Security related network functions. 3GPP, December 2012.
[12] D. Dolev and A. Yao. On the security of public key protocols. IEEE Transactions on Information Theory, 29(2): 198—208, March 1983.
[13] M. Blaze, W. Diffe, R. Rivest, B. Schneier, T. Shimomura, E. Thompson, and M. Wiener. Minimal key lengths for symmetric ciphers to provide adequate commercial security. Report of Ad Hoc Panel of Cryptographers and Computer Scientists, January 1996. Available from `http://www.crypto.com/papers/`.
[14] N. Smart (Ed.). ECRYPT II Yearly Report on Algorithms and Keysizes (2010–2011) ECRYPT II NoE, ICT-2007-216676, Deliverable D.SPA.17, Rev.1, June 2011.
[15] R. Anderson. Why cryptosystems fail. In Proceedings of the 1st ACM Conference on Computer and Communications Security (CCS'93). ACM Press, 1993.
[16] B. Schneier. Beyond Fear: Thinking Sensibly about Security in an Uncertain World. Springer, 2003.
[17] D. Florêncio and C. Herley. Where do all the attacks go? Microsoft Research, Technical Report 2011-74, 2011. Available from `http://research.microsoft.com/pubs/149885/WhereDoAllTheAttacksGo.pdf`.
[18] D. Pavlovic. Gaming security by obscurity. In Proceedings of the 2011 Workshop on New Security Paradigms Workshop (NSPW 2011), 2011.
[19] The rising cost of identity theft for consumer. In Bucks Blog, New York Times, 2011/02/09, 2011.
[20] Fighting fraud together; A strategic plan to reduce fraud, 12 October 2011. Home Office, UK, 2011.
[21] S. Kent and K. Seo. RFC 4301: Security architecture for the Internet protocol. IETF RFC 4301, December 2005.
[22] G. M. Køien and V. A. Oleshchuk. Spatio-temporal exposure control: An investigation of spatial home control and location privacy preserving issues. In Proceedings of the 14th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2003), Beijing, China, 7–10 September, pp 2760–2764. IEEE Press, 2003.

[23] D. Erickson et al. A demonstration of virtual machine mobility in an OpenFlow Network. In Proceedings of ACM SIGCOMM'08, 17–22 August 2008.
[24] E. Aïmeur and D. Schönfeld. The ultimate invasion of privacy: Identity theft. In Proceedings of the Ninth Annual International Conference on Privacy, Security and Trust (PST11), Monteral, Canada, August 2011.
[25] Office of the Information and Privacy Commissioner of Ontario. In Privacy by Design: Time to Take Control. `www.privacybydesign.ca`, Ontario, Canada, January 2011.
[26] G. M. Køien. Privacy enhanced device access. In Proceedings of MobiSec 2011, Aalborg, Denmark, May 2011.
[27] G. M. Køien. Privacy enhanced cellular access security. In Proceedings of the 2005 ACM Workshop on Wireless Security, pp. 57–66, Cologne, Germany, September 2005.
[28] G. M. Køien and V. A. Oleshchuk. Location privacy for cellular systems; Analysis and solution. PET 2005, Cavtat, Croatia, LNCS, Vol. 3856. Springer, 2005.
[29] V. A. Oleshchuk and G. M. Køien. Security and privacy in the cloud; A long-term view. In Proceedings of Wireless VITAE, pp. 1–5, 2011.
[30] ISO, ISO/IEC 11990-1 Information technology – Trusted Platform Module – Part 1: Overview, 2009.
[31] N. Ulltveit-Moe, V. A. Oleshchuk, and G. M. Køien. Location-aware mobile intrusion detection with enhanced privacy in a 5G context. Wireless Personal Communincations, 57(3), 2010.
[32] A. Schwartz. Privacy and security identity management and privacy: A rare opportunity to get it right. Communications of the ACM, 54(6): 22–25, June 2011.
[33] C. Boyd and A. Mathuria. Protocols for Authentication and Key Establishment. Springer Verlag, 2003.
[34] L. Rajbhandari and E. Snekkenes. Intended actions: Risk is conflicting incentives. In Proceedings of the 15th International Information Security Conference (ISC 2012), LNCS, Vol. 7483, pp. 370–386. Springer, 2012.

Biography

Geir M. Køien is an associate professor at the University of Agder, Norway. His primary research interests are system security, personal privacy and cellular access secuirty. He has previously worked for Telenor R&D, where he was a delegate to the 3GPP SA3 security work group for 10 years. Currently he also holds an adjunct position with the Norwegian Post and Telecommunications Authority as a senior advisor on cellular security.