Fast authentication and trust-based access control in heterogeneous wireless networks

Maryna Komarova

École Doctorale
d'Informatique,
Télécommunications
et Électronique de Paris

Thèse

Présenté pour obtenir le grade de docteur
de Telecom-ParisTech

Spécialité : Informatique et Réseaux

par
Maryna Komarova

Sujet:

AUTHENTIFICATION RAPIDE ET CONTROLE D'ACCES BASE SUR LA CONFIANCE DANS LES RESEAUX SANS FIL HETEROGENES

Soutenue le 19 mai 2008 devant le jury composé de :

Fabio Martinelli	Istituto di Informatica e Telematica	Rapporteur
Isabelle Chrisment	Université Henri Poincaré (Nancy 1)	Rapporteur
Xavier Lagrange	TELECOM Bretagne	Examinateur
Jean Leneutre	TELECOM ParisTech	Examinateur
Michel Riguidel	TELECOM ParisTech	Directeur de thèse

École Doctorale
d'Informatique,
Télécommunications
et Électronique de Paris

PhD Thesis

Presented to obtain the degree of PhD
at the Computer Science and Networks Department
of the Telecom-ParisTech

by
Maryna Komarova

Subject:

FAST AUTHENTICATION AND TRUST-BASED ACCESS CONTROL IN HETEROGENEOUS WIRELESS NETWORKS

Jury:

Fabio Martinelli	Istituto di Informatica e Telematica	Reporter
Isabelle Chrisment	Henri Poincaré University (Nancy 1)	Reporter
Xavier Lagrange	TELECOM Bretagne	Examiner
Jean Leneutre	TELECOM ParisTech	Examiner
Michel Riguidel	TELECOM ParisTech	Thesis supervisor

ACKNOWLEDGEMENTS

To my family

Моей семье

I have an opportunity to express my gratitude to all people who have contributed to this work, supported and accompanied me during my thesis preparation.

First of all, I would like to express my sincere gratitude to my thesis supervisor, Professor Michel Riguidel for the guidance, continuous encouragement and patience. Many thanks for his suggestions, comments, attentive reading, improvement of the technical quality of my work and interesting and fruitful discussions. I have learned a lot from him.

I wish to deliver my gratitude to the honourable jury members, Professor Isabelle Chrisment (LORIA, Henri Poincaré University), Doctor Fabio Martinelli (Instituto di Informatica e Telematica), Professor Xavier Lagrange (ENST Bretagne) and Doctor Jean Leneutre (Telecom-ParisTech) for accepting to be a part of the thesis committee and for their constructive review of my thesis. Comments and suggestions provided by them have helped me to improve the quality of this work.

I thank the members of the Computer science and networks department (INFRES) of Telecom-ParisTech and Ecole Doctorale d'Informatique, Télécommunications et Electronique (EDITE) for excellent working conditions creation, professional and human support and friendly company. My particular thanks for Dr. Artur Hecker for his valuable input and encouragements. I greatly appreciate help and advices provided by Dr. Jean Leneutre and Prof. Philippe Godlewski.

During my thesis preparation I had a great pleasure to work with members of IEEE 802.21 Working group. They helped to enrich my vision of mobility and related security problems. I wish to thank Security Study group members, notably Yoshihiro Ohba, Michael Williams and Subir Das for discussions and comments.

I would like to thank my friends for support, motivation, interesting and rich spare time that they offer me.

And of course I am grateful to my husband Ivan for his support, understanding and simply for being here. This work would not be done without him.

RÉSUMÉ

1 INTRODUCTION GENERALE

Durant les dernières années, le développement intensif des technologies de communication sans fils a conduit à la création d'un réseau d'accès global et omniprésent. Les services offerts aux utilisateurs varient des services simples et traditionnels comme l'échange de données et de courrier aux applications en temps réel comme la Voix sur IP ou la vidéo à la demande. L'univers numérique actuel est caractérisé par la coexistence des réseaux d'accès, une multiplicité de technologies, une diversité de services, des opérateurs multiples, des architectures différentes et des protocoles distincts. L'évolution des terminaux portables, la non-nécessité d'une reconnexion manuelle à chaque réseau sans fil visité ont favorisé la mobilité des utilisateurs. Dans ce contexte, la possibilité d'accéder aux services indépendamment du réseau d'accès et du terminal a été fortement sollicitée. De plus, le changement du réseau d'attachement doit être transparent pour l'utilisateur et ne doit pas provoquer une dégradation de la qualité de service ou la rupture de la session.

Une sécurité forte est très importante dans un environnement mobile. En plus des risques connus sur des réseaux filaires, de nouvelles vulnérabilités propres aux communications sans fil apparaissent.

Dans cette thèse, la situation actuelle et les évolutions dans le développement des technologies sans fil sont présentées. Les problèmes liés aux solutions de sécurité implémentées sur des technologies différentes et, en particulier la possibilité de les intégrer a été analysée. Les solutions de sécurité ont toujours des modèles de confiance sous-jacents. Dans un univers communicant dynamique, des procédures permettant d'établir la confiance entre les parties, doivent être définies. Les mécanismes de sécurité du support de la mobilité doivent être flexibles, adaptables aux conditions, indépendants de l'opérateur, de l'infrastructure et de la technologie sous-jacente. Les défis d'une mobilité ubiquitaire sécurisée ont été identifiés et les solutions existantes, avec leurs avantages et défauts, ont été analysées.

Ensuite, nous examinons la possibilité de réduire la latence d'authentification dans le cas où l'authentification pour l'accès réseau et l'authentification pour le service sont des processus séparés. Pour ce faire, en exécutant une seule authentification, l'utilisateur a l'autorisation d'accéder au réseau et au service. L'approche d'authentification ainsi proposée combine les opérations de 802.1X et PANA.

Puis, nous introduisons le protocole d'authentification rapide (Fast re-Authentication Protocol) pour la transition inter – domaine. Ce protocole a pour

but de réduire le délai d'authentification de l'utilisateur mobile dans un réseau où cet utilisateur n'est pas abonné. L'approche proposée élimine la nécessité de communication entre le réseau cible et le réseau d'origine de l'utilisateur afin de vérifier son identité. Nous développons le protocole d'authentification rapide par la proposition d'un schéma pour la distribution de tickets d'authentification. Cette méthode permet de diminuer le nombre des tickets générés et envoyés et, par conséquent, la surcharge et le délai de la phase d'acquisition des tickets du protocole.

Les résultats numériques obtenus en cours d'expérience sur un banc d'essai et une série de simulations ont montré que l'approche proposée améliore les paramètres de handover tels que la latence d'authentification et le coût de la signalisation.

Afin de rendre le contrôle d'accès plus flexible et adaptable à l'environnement dynamique, nous introduisons le mécanisme de contrôle d'accès basé sur la confiance. Ce modèle repose sur une prise en compte de l'expérience d'interaction, les recommandations et les réputations des fournisseurs de recommandations. L'importance de chaque composant évolue avec le temps.

L'expérience d'interaction ou la confiance basée sur l'observation, est construite à l'aide d'un modèle de confiance ajustable. Ce modèle permet d'adapter les droits d'accès d'un utilisateur selon son comportement pendant les interactions passées. De plus, les mécanismes pour adapter les politiques d'accès au niveau du risque actuel sont proposés. Dans le contexte de ce modèle nous proposons une solution pour retenir l'historique de chaque utilisateur et pour modifier l'importance de cet historique avec le temps. Cette solution a été validée par les simulations.

Finalement, nous envisageons l'intégration de l'authentification rapide et le contrôle d'accès basé sur confiance dans une approche générique.

2 SITUATION ACTUELLE DANS L'UNIVERS DE COMMUNICATION SANS FIL

Les technologies sans fil permettent la connectivité entre un utilisateur et le réseau filaire de communication globale tel que Internet ou le réseau téléphonique. Les réseaux sans fil sont devenus populaires dans l'activité professionnelle et quotidienne. Les technologies sans fil offrent une grande variété de capacités orientées vers des buts différents. Ces avantages sont nombreux : les connexions possibles là où les connexions filaires ne le sont pas, l'installation des réseaux est plus facile et moins chère et ce type d'accès offre surtout la mobilité aux utilisateurs. L'ensemble des réseaux sans fil évolue vers un réseau proposant plusieurs services, offrant l'accès par plusieurs technologies, géré par de multiples opérateurs et supportant la mobilité pour l'utilisateur. Les caractéristiques principales du futur

réseau omniprésent sont la mobilité, la transparence et l'ubiquité. Les technologies sans fil ont été développées pour les conditions d'usage différent, et il n'existe pas de technologie universelle parfaite. Certaines proposent un meilleur support pour la transmission de la voix alors que les autres sont plus adaptées pour la transmission des données. Une classification des réseaux sans fil peut être basée sur la technologie de transmission ou sur la taille du réseau. Selon ce dernier critère, il y a quatre groupes de réseaux sans fil de la portée la plus courte à la plus grande : les WPAN (Wireless Personal Area Networks) dont la norme IEEE 802.15 est la plus utilisée, les WLAN (Wireless Local Area Networks) représentés par la norme IEEE 802.11a/b/g, les WMAN (Wireless Metropolitan Area Networks) parmi lesquels IEEE 802.16e se développe rapidement et WWAN (Wireless Wide Area Networks) représenté par la famille UMTS.

Le terme « mobilité » est associé à la possibilité de l'utilisateur d'accéder à des services de télécommunication quel que soient le lieu et le terminal. La continuité de la session active pendant le changement de point d'attachement est sous-entendue. Selon l'objet qui se déplace, la mobilité peut être classée comme la mobilité du terminal, de la session, de l'utilisateur, de l'application et du code. Dans cette thèse, nous nous concentrons sur l'analyse de mobilité des terminaux. Les réseaux cellulaires et ceux de la norme 802.16e ont été conçus avec le support natif de la gestion de mobilité inter - cellule. Pour assurer la mobilité dans d'autres types de réseaux, des entités et protocoles additionnels sont nécessaires. Les mécanismes de gestion de la mobilité locale sont spécifiques pour chaque technologie. Afin d'assurer la mobilité globale et indépendante de la technologie deux protocoles sont proposés : IP Mobile qui gère la mobilité de terminaux et SIP supportant la mobilité de la session.

La mobilité inter - domaine est constituée du *handover* (Transition) et du *roaming* (Itinérance). Le Handover est un passage sans coupure d'une cellule à une autre adjacente (peu importe si ces cellules appartiennent à des réseaux différents ou au même réseau). Le handover peut avoir lieu au cours de la session. Les caractéristiques du processus de handover sont la latence, la perte de paquets et la variation, définis pour estimer l'influence sur la qualité de la communication. La relation de roaming est un accord entre les opérateurs permettant aux abonnés d'un opérateur à accéder aux ressources de l'autre opérateur.

La situation actuelle dans le monde sans fil conduit à l'apparition des nouvelles demandes de la part des utilisateurs mobiles. La mobilité doit être ubiquitaire; l'utilisateur mobile doit pouvoir se déplacer entre les réseaux gérés par des autorités différentes, éventuellement en changeant la technologie d'accès. Les services offerts par les opérateurs ont évolué de la voix analogique à la voix sur IP. Les applications en temps réel exécutées sur les terminaux mobiles sont très sensibles à la latence de transition entre les points d'attachement. Finalement, l'authentification mutuelle doit être exécutée entre l'utilisateur mobile et le réseau visité.

3 DE LA CONFIANCE A LA SECURITE

Une large utilisation des télécommunications sans fil implique des nouveaux besoins et défis de sécurité en plus des risques connus sur les réseaux fixes. Certaines failles de sécurité sont communes pour les réseaux sans fil de tous types. Afin de protéger l'infrastructure et les communications, chaque technologie possède ses propres mécanismes de sécurité.

La mobilité apporte de nouveaux défis de sécurité. Dans l'univers numérique traditionnel les objectifs principaux de sécurité sont la confidentialité, l'intégrité, la disponibilité et la sûreté de fonctionnement. Cet univers peut être représenté comme un environnement statique, caractérisé par des frontières, les failles, les listes des participants plus ou moins définis et les relations de confiance statiques et pour la plupart directes. Les communications mobiles introduisent un aspect dynamique. En plus de la confidentialité des données, la confidentialité de l'emplacement, du trafic et de l'identité doit être adressée. L'utilisation de la cryptographie dans la transmission devient nécessaire mais difficile à mettre en œuvre à cause de capacités limitées des terminaux mobiles. De plus, les mécanismes de sécurité sont coûteux en termes de temps et de ressources consommées.

Les modèles de sécurité sont en relation avec les modèles de confiance. Dans l'environnement dynamique avec des marges mal définies les mécanismes de sécurité traditionnels sont en général utilisés. Ces modèles de confiance sont statiques et ils reflètent les relations entre les entités communicantes, au moment de la création des relations. Nous avons étudié les relations entre la sécurité et la confiance ainsi que les modèles de confiance utilisés dans la commerce électronique et dans les réseaux pair à pair. Nous avons analysé les caractéristiques de la confiance afin de mieux comprendre comment ces relations de confiance peuvent être construites en façon dynamique.

Les relations de confiance peuvent être établies parmi des fournisseurs d'accès au réseau et des fournisseurs de services en formant un accord bilatéral ou une fédération multilatérale. L'utilisateur abonné à une autorité peut ainsi être servi par une autre, avec laquelle il n'a pas de relations de confiance directe. Puisque les relations de confiance ne sont pas transitives, l'infrastructure de confiance et des mécanismes appropriés sont nécessaires pour établir la confiance entre un utilisateur mobile et un réseau visité. L'authentification mutuelle est un exemple de ces mécanismes. Il existe deux approches fondamentales pour vérifier l'identité d'utilisateur et celle du réseau visité. La première d'entre elles exige la communication entre le réseau visité et le réseau d'origine de l'utilisateur. Cet échange provoque un délai imprévisible et difficile à réduire. Le deuxième groupe des méthodes est basé sur la cryptographie asymétrique. L'utilisation de certificats à clé publique peut éliminer la nécessité d'une communication inter-domaine durant

le processus de l'authentification. Les mécanismes de la cryptographie asymétrique et du traitement des certificats sont coûteux en termes de temps et des ressources consommées.

Pour atteindre les objectifs de sécurité du réseau, des mécanismes de contrôle d'accès doivent être déployés. Les mécanismes de contrôle d'accès permettent aux utilisateurs autorisés d'accéder aux ressources du système. Les autorisations sont basées sur la définition des permissions appelées les politiques d'accès. Une politique d'accès est une liste de contraintes spécifiques ayant pour but de protéger un système donné. Nous avons présenté le suivi de l'évolution des mécanismes du contrôle d'accès et analysé les inconvénients sur l'application dans un système ouvert et hétérogène.

4 VERS L'INTERCONNEXION DES RESEAUX UBIQUITAIRE : NOTRE VISION

L'univers numérique moderne est hétérogène dans plusieurs sens. Les nombreux services basés sur IP sont offerts aux utilisateurs qui sont abonnés aux nombreux fournisseurs de services et qui ont des rôles divers dépendant de la situation. Ces utilisateurs sont équipés d'appareils portables avec des capacités différentes. Les utilisateurs peuvent ainsi accéder à une large gamme de services via de nombreux réseaux d'accès gérés par des opérateurs différents. Le cadre limité de chaque technologie d'accès permet à l'utilisateur de choisir son réseau avec les caractéristiques appropriées à son besoin courant. Des technologies différentes coexistent dans la même région géographique. Cet univers hétérogène nécessite des nouveaux paradigmes et approches pour gérer les types de mobilité diverses et pour fournir aux utilisateurs des services adaptés au mode d'accès et au terminal. Dans cette thèse nous avons présenté notre vision d'évolution des communications sans fil. Les défis principaux pour la mobilité ubiquitaire et sécurisée sont l'extensibilité, l'interopérabilité, la garantie de QdS durant l'exécution du handover et, bien sûr, la sécurité.

Les approches de la gestion de la mobilité ont montré une bonne performance mais le délai du handover est toujours affecté par la signalisation liée à la sécurité. Les besoins simultanés de la sécurité forte et de la minimisation de la latence du handover s'avère un problème complexe.

L'implémentation des solutions de sécurité dans l'environnement mobile a des nombreuses contraintes. D'abord, les procédures d'authentification et d'établissement de confidentialité ne doivent pas augmenter la latence du handover. Les protocoles d'authentification ne doivent pas utiliser des procédures de calcul lourdes à cause de la capacité limitée des batteries des terminaux portables et du

temps limité. Les problèmes liés la sécurité de la mobilité peuvent être classés comme techniques et non-techniques.

Nous avons analysé les problèmes liés à la mobilité sécurisée. Apres avoir décrit notre vision d'évolution des réseaux sans fil vers ces convergences, nous avons identifié les failles de sécurité pour les processus d'handover inter-domaine et les défis principaux. Des efforts sont réalisés par les organisations tant industrielles, académiques que normatives dans le but de surmonter l'hétérogénéité des solutions de sécurité et de la gestion de mobilité. Dans cette thèse l'analyse des solutions pour assurer la mobilité ubiquitaire et sécurisée est présentée.

- ✓ **Problème de la sélection du réseau candidat** (cible) : Pour changer de point d'attachement, l'utilisateur doit obtenir l'information sur les réseaux disponibles et choisir la cible; s'authentifier auprès du réseau choisi; établir les clefs de chiffrement et mettre à jour sa position auprès du réseau d'origine et du nœud correspondant. Les protocoles d'authentification les plus utilisés, ont été développés sans prendre en considération la mobilité des utilisateurs. L'implémentation des solutions pour l'authentification rapide développées actuellement est limitée pour la technologie d'accès particulière (par exemple, GSM ou 802.11) ou pour les réseaux d'accès gérés par un opérateur, comme Unlicensed Mobile Access (UMA) proposé par 3GPP.
- ✓ **Maintien du niveau de la sécurité et correspondance des mécanismes de la sécurité** : L'utilisateur mobile se connecte aux réseaux différents en espérant que ses données et son identité seront protégées. La plupart des normes permettent à l'utilisateur de négocier les paramètres de sécurité avant l'association. Par contre, il n'y a pas de réponse à la question de savoir comment sont comparés et mis en correspondance les mécanismes de sécurité implémentés dans des réseaux différents.
- ✓ **Constitution dynamique des relations de confiance** : La mobilité de l'utilisateur est limitée par les relations entre les partenaires de son réseau d'origine, des autres réseaux, où il est abonné et des réseaux libres. Il n'y a pas en général de relations de confiance implicites entre un utilisateur mobile et un réseau visité, ils doivent donc établir les relations de confiance de manière dynamique. Ce problème est étroitement lié à celui de l'authentification rapide.
- ✓ **Authentification rapide réciproque** : Le défi principal est lié à l'authentification entre l'utilisateur et le réseau visité car tous les mécanismes de confidentialité et d'intégrité dépendent du résultat de l'authentification. Pour protéger le réseau de l'accès non autorisé, l'authentification doit être effectuée de préférence sur la couche de transmission. Pour assurer les transitions inter-technologies, les mécanismes

de sécurité doivent être indépendants de la technologie sous-jacente. Il a été montré que l'authentification est responsable la part majeure de la latence du handover. Plusieurs solutions pour l'authentification rapide ont été proposées mais les applications sont limitées à un domaine administratif ou à une seule technologie de transmission.

- ✓ **Contrôle d'accès dans un environnement ouvert** : La mobilité implique des problèmes liés à l'autorisation. Le domaine visité doit déterminer quelles sont les paramètres d'autorisation du visiteur. Le contrôle d'accès traditionnel est convenable pour un environnement où les utilisateurs sont connus à l'avance et où ses droits et restrictions sont prédéterminés. Ce type d'environnement est illustré par les réseaux domestiques ou universitaires. Le cas général de mobilité ubiquitaire sous-entend un manque de confiance parmi les participants. Plusieurs modèles de contrôle d'accès basé sur la confiance ont été proposés. La plupart d'entre eux ne fixe pas la correspondance directe et transparente entre le niveau du risque dans l'environnement, les politiques d'accès et les paramètres du modèle de confiance.

Les communications numériques modernes exigent des solutions de sécurité plus flexibles et autonomes que ceux que les mécanismes traditionnels de sécurité offrent actuellement.

5 CONTRIBUTIONS

Les contributions présentées dans cette thèse traitent des questions de réduction du temps d'authentification inter-domaine, d'optimisation de la signalisation liée à la sécurité et du contrôle d'accès dans un environnement ouvert.

5.1 L'authentification composée dans un réseau sans fil visité

Tout d'abord, nous avons étudié la possibilité de réduction des ressources d'authentification dans le cas où l'authentification au réseau d'accès est séparée de l'authentification au service, par exemple, l'accès à l'Internet. Nous proposons une méthode composée pour l'authentification de l'utilisateur dans un réseau sans fil public. Cette approche a pour but de minimiser le risque des attaques sur les noeuds réseau internes lancées par les utilisateurs non authentifiés, et de fournir les clés de chiffrement entre le nœud mobile et le point d'accès et de réduire ainsi la latence totale du processus d'authentification. Après avoir exécuté l'authentification combinée, l'utilisateur obtient l'accès aux éléments de l'infrastructure du réseau et au service.

Le protocole développé est destiné à l'implémentation dans un réseau qui utilise PANA (Protocol for Carrying Authentication for Network Access) et 802.1X avec EAP comme protocoles d'authentification. L'utilisation de PANA exige la présence de la sécurité de communication en dessous de IP. L'idée principale est de combiner les opérations de deux protocoles dans une seule.

La norme de PANA définit le routeur d'accès comme « l'authenticateur ». Ceci facilite la pré-authentification inter-domaine car le routeur d'accès peut communiquer avec les entités extérieures. Par contre, ce protocole ne fournit pas de moyens pour sécuriser les communications sur la couche de transmission.

Dans l'approche proposée, PANA et 802.11i partagent les tâches : le premier est utilisé pour l'authentification de l'utilisateur et le deuxième est utilisé pour la négociation des clés de chiffrement et pour le contrôle d'accès général au réseau. La point d'accès joue un double rôle : il est aperçu par l'utilisateur comme « l'authenticateur » et comme le client par le PANA Agent. Durant le processus d'authentification, le point d'accès transmet les messages venant de l'utilisateur au PANA Agent qui communique avec le Serveur d'Authentification. Dans le cas où l'authentification réussit, le terminal de l'utilisateur est capable d'obtenir l'adresse IP interne du réseau. Pour indiquer que le client change d'adresse, le point d'accès communique à PANA Agent l'option « Post-PANA Address Configuration ».

La combinaison des opérations améliore la performance du handover, en particulier, la méthode proposée permet d'éviter l'acquisition double d'adresse IP par le terminal d'utilisateur et de diminuer le temps d'authentification jusqu'à 100 ms par rapport à des authentifications consécutives. La performance de l'approche a été analysée à l'aide des simulations en utilisant Omnet++.

5.2 L'authentification rapide pour la mobilité inter – domaine

Le Fast re-Authentication Protocol (FAP) protocole est proposé pour l'authentification inter - domaine. Ce protocole est développé afin de diminuer le délai d'authentification de l'utilisateur mobile dans un domaine visité. L'approche élimine la nécessité de la communication entre le domaine visité et le domaine d'origine d'utilisateur pour la vérification d'identité de ce dernier. L'authentification proposée est basée sur l'utilisation de la structure légère et de courte durée que l'on a appelée « ticket d'authentification ». Le protocole proposé est indépendant de la nature d'associations de sécurité entre les réseaux différents.

La solution proposée se compose de deux protocoles :

- ✓ L'acquisition des tickets;
- ✓ L'authentification rapide.

Les hypothèses avancées sont:

- ✓ Le réseau cible a des accords de roaming, soit avec le réseau d'origine, soit avec le réseau servant;
- ✓ Les partenaires partagent les clés;
- ✓ Le terminal d'utilisateur a déjà accompli l'authentification initiale en utilisant n'importe quelle méthode.

Fast re-Authentication Protocol spécifie les communications entre le serveur FAP du côté réseau et le client FAP du côté utilisateur. La phase d'acquisition des tickets est conçue pour fournir aux utilisateurs des preuves d'identité pour l'authentification rapide postérieure. Le serveur FAP génère les tickets après l'authentification d'utilisateur qui réussit et il ne fait pas de renouvellement ou de révocation des tickets. Si le ticket expire, le nouveau peut être généré sur la demande de l'utilisateur. Le ticket d'authentification:

- ✓ Contient le secret fourni par le réseau partenaire du réseau candidat,
- ✓ Est créé après chaque authentification forte qui réussit,
- ✓ Peut être vérifié seulement par son créateur ou destinataire,
- ✓ Permet de réduire le temps d'obtention d'accès au réseau,
- ✓ Peut être créé soit par le réseau d'origine, soit par le réseau servant selon les politiques.

Le ticket est lié aux réseaux cible et émetteur par la clé partagée entre eux. Il est aussi lié à l'utilisateur par son pseudonyme et le résultat de l'authentification précédente. Le ticket se compose de deux parties : chiffrée, déchiffrable seulement par l'émetteur et la cible, et la partie en clair, permettant à l'utilisateur de connaître le nom du destinataire et le temps d'expiration. L'ensemble de l'information est signé afin de prévenir une modification par l'utilisateur ou par un attaquant.

Avant de commencer le processus d'authentification, l'utilisateur possède un ticket d'authentification correspondant au réseau candidat. Si le réseau candidat est dans la liste des partenaires du réseau d'origine d'utilisateur et si le ticket est expiré ou absent, l'utilisateur peut lancer la demande de ticket à son réseau d'origine. L'échange d'authentification se compose de quatre messages. Le client commence la communication avec la requête d'accès qui contient le ticket et deux nombres aléatoires. Le serveur FAP répond avec un Challenge qui contient une fonction d'un nombre aléatoire reçu et un nombre généré, chiffrée avec une clé dérivée de l'information du ticket, et un autre nombre aléatoire. Le client déchiffre le message, dérive le matériel pour la génération des clés de chiffrement et envoie le Response message au FAP serveur. Si le contenu de la réponse correspond au Challenge envoyé, le serveur répond avec un Success message, sinon avec un Failure message.

Le fonctionnement du protocole est indépendant de la nature des associations de sécurité entre les réseaux partenaires. La méthode proposée permet de réduire le temps d'obtention d'accès au réseau et de générer les clés du chiffrement ultérieur.

Pour illustrer l'opération de FAP nous avons choisi la technologie IEEE 802.11. Pour estimer le délai de la phase d'authentification de FAP nous l'avons implémenté sur un banc d'essai. Nous avons modifié le logiciel du serveur (RADIUS) et du client (Xsupplicant) en ajoutant une nouvelle méthode EAP, appelé EAP-FAP. Le délai d'authentification moyen pour 100 expériences a été 30.59 ms. Le délai d'authentification pour EAP-TTLS avec MD5 sous les mêmes conditions a été 85.33 ms. Le protocole proposé a donc montré une latence d'authentification inférieure à la durée d'authentification des protocoles standards.

5.3 Schéma pour la distribution de tickets efficace

La distribution efficace du matériel d'authentification est nécessaire pour assurer la mobilité inter – domaine. Afin de satisfaire cette demande, une approche pour l'optimisation de la distribution des tickets d'authentification est proposée.

Le protocole d'authentification rapide (FAP) décrit dans la section précédente réduit le temps d'authentification, mais les utilisateurs qui changent de réseaux souvent peuvent provoquer une surcharge de trafic entre ces utilisateurs et leurs réseaux d'origine.

Pour minimiser le nombre des tickets envoyés à chaque abonné, l'utilisation de table de voisins, dont chaque ligne correspond au partenaire de roaming et contient ses voisins, est proposée. Cette table est située sur le serveur d'authentification de chaque réseau. Le réseau ajoute un voisin sur la requête de l'utilisateur qui change de réseau d'attachement.

Pour créer la table des voisins, deux modes d'opération sont proposée pour FAP : proactif et réactif. Dans le mode réactif le terminal mobile choisit le réseau cible. Puis le terminal envoie le ticket requête à son réseau d'origine ; ce dernier ajoute un voisin dans la table de partenaires et répond au client avec le ticket correspondant. Quand le protocole fonctionne dans le mode proactif, le terminal de l'utilisateur est attaché à un réseau visité et il ne connaît pas les destinations potentielles. Juste après l'authentification dans ce réseau l'utilisateur reçoit les tickets, seulement pour les voisins du réseau d'attachement actuel. Le mode réactif permet à l'utilisateur d'obtenir le ticket d'authentification après avoir choisi le réseau cible.

Les résultats numériques obtenus au cours des simulations ont démontré que l'approche proposée peut améliorer les paramètres de transition inter – domaine tels que la latence d'authentification et le coût de la signalisation.

L'acquisition des tickets d'authentification

- ✓ Diminue le volume de trafic entre l'utilisateur mobile et son réseau d'origine et la charge du serveur d'authentification;
- ✓ Permet de combiner la mise à jour d'emplacement de l'utilisateur avec la construction de la table des voisins.

5.4 Méthode de contrôle d'accès basée sur confiance

Les utilisateurs mobiles peuvent s'authentifier dans les réseaux de types divers. Un utilisateur peut se déplacer parmi les réseaux partenaires de son réseau d'origine, soit d'autres réseaux où il est abonné ou bien des réseaux à but non lucratif. Si les accords de roaming sont établis entre les autorités, l'abonné de l'un peut être servi par un autre. L'utilisateur recommandé par l'autorité partenaire peut montrer un comportement « positif » ou « négatif ». Les protocoles d'authentification et de contrôle d'accès utilisent des modèles de confiance statique, qui n'ont pas de capacité à réagir au comportement des clients. Les systèmes de commerce électronique utilisent les modèles dynamiques de confiance.

Ce travail a pour but de donner au fournisseur des services ou des ressources l'opportunité d'évaluer la fiabilité d'un client, de réagir au comportement des clients par adaptation des politiques d'accès au niveau du risque actuel. Les privilèges d'un utilisateur sont définis par son comportement précédent, des recommandations de tiers de confiance et de la situation courante. Les relations de confiance peuvent être établies si au moins une source de confiance est disponible. L'approche fonctionne en trois étapes : le client s'authentifie avec le réseau, le réseau estime la valeur de confiance correspondante et, finalement, les politiques d'accès sont mises en correspondance avec cette valeur de confiance.

Nous considérons que la valeur de confiance à un client est formée de la valeur d'expérience (ou la confiance fondée sur les observations), de la recommandation d'un tiers de confiance et de la réputation de l'entité qui a fourni le recommandation. La réputation sert de poids de fiabilité de la recommandation. Dans notre modèle, l'influence de chaque source de confiance change avec le temps. Si un fournisseur de service n'a pas assez d'information sur un client inconnu, la confiance à ce client est fondée sur la recommandation, mais si un client est « bien connu » par le fournisseur, la recommandation a moins d'importance.

Ce mécanisme pour le contrôle d'accès adresse les défis présentés par l'environnement mobile et ouvert. Nous avons ajouté l'aspect dynamique à la gestion des relations de confiance entre les entités qui fournissent des services. Dans le cas où les relations de roaming existent entre deux réseaux, chacun d'entre eux

peut évaluer la confiance en l'autre basée sur l'observation de l'activité des utilisateurs recommandés.

Les scénarios considérés incluent les interactions entre les pairs dans un réseau d'overlay, entre l'utilisateur et les services Web et entre l'utilisateur et le réseau d'accès.

5.5 Définition du modèle de confiance ajustable

Pour un réseau public, il est souhaitable de donner l'accès au client en prenant en considération son comportement passé. Le modèle de confiance proposé doit permettre au réseau de s'adapter à l'environnement qui change, par l'adaptation des politiques de sécurité et du contrôle d'accès.

On suppose que le réseau est capable d'observer le comportement des utilisateurs. La formalisation de la confiance sert à mieux prendre en compte l'historique du comportement des utilisateurs afin d'estimer le risque que chaque utilisateur représente pour le réseau, de limiter l'accès pour les utilisateurs non fiables et de favoriser les « bons » utilisateurs.

L'expérience fondée sur l'observation est une source de confiance plus sûre. Le modèle ajustable de confiance est développé pour l'amélioration du contrôle d'accès. Ce modèle permet au réseau de servir seulement les clients, qui sont reconnus comme les «non- malveillants ». Le réseau peut fournir plusieurs services et pour accéder à chacun de ces services, le client doit atteindre un certain niveau de confiance. La valeur de la confiance à un client est calculée en prenant en considération les interactions précédentes entre ce client et le réseau.

Sur chaque requête d'accès d'un client, le réseau calcule la valeur de confiance en utilisant un modèle linéaire. Si le client montre un « bon » comportement, la valeur de la confiance croît jusqu'au niveau maximal avec l'augmentation du nombre des visites. L'utilisateur inconnu est considéré comme un « bon » utilisateur et peut accéder aux services de base. La valeur de la confiance est définie par le nombre des expériences positives que le réseau a eu avec le client et deux paramètres, nommés l'optimisme et la tendance. La vitesse de croissance de la valeur de la confiance est fixée par le paramètre *optimisme*. La *tendance* est la valeur de confiance maximale, que l'utilisateur peut atteindre. Ces paramètres changent en fonction du nombre d'expériences négatives.

Dans le contexte de ce modèle nous avons proposé une approche pour conserver l'historique des interactions avec les clients et pour modifier l'importance de cet historique avec le temps. Pendant la création d'un modèle de confiance, il faut tenir compte de plusieurs limitations, notamment de la limitation de la taille de mémoire, que le système peut consacrer à l'historique de comportement des clients, et du facteur temporel. Le facteur temporel est très important parce que les

événements plus récents ont plus d'influence sur la décision finale pour ce qui concerne la fiabilité d'un utilisateur. Pourtant, l'information sur le comportement passé doit aussi être prise en considération. Expérience ancienne ne signifie pas expérience obsolète ou inutile. Au lieu d'utiliser des fenêtres de mémoire, la mémoire fading ou des poids d'oubli, proposés dans la littérature, nous avons proposé de conserver l'historique de l'ensemble des variables scalaires (nombre des expériences positives, nombre des expériences négatives, nombre de fois que le client a perdu la confiance, le niveau de confiance maximale, temps d'oubli). Dans le modèle proposé le client perd la confiance et le droit d'accéder le réseau après avoir montré un certain nombre de fois un comportement négatif. Cependant, ce client peut être « pardonné » et regagner les droits d'accès.

Les paramètres utilisés dans le modèle proposé dépendent de politiques d'accès qui peuvent changer selon le niveau du risque actuel. La modification des politiques d'accès est faite d'une manière autonome et automatique. Nous avons proposé un modèle de confiance caractérisé par la correspondance claire entre les politiques d'accès, les paramètres du modèle et la valeur de confiance obtenue.

Dans le cadre de ce travail nous avons proposé une méthode simple permettant à l'utilisateur de choisir le réseau candidat, fondée sur la réputation des réseaux voisins.

6 CONCLUSIONS ET PERSPECTIVES

6.1 Conclusions

Dans cette thèse nous avons examiné les défis liés à la mobilité ubiquitaire sécurisée dans un environnement hétérogène. Ces problèmes ont été analysés du point de vue d'un utilisateur mobile et d'un réseau d'un fournisseur de services. L'hétérogénéité des technologies d'accès sans fil crée des exigences particulières pour la gestion de la mobilité et pour la conception de mécanismes de sécurité. D'un coté, ces mécanismes doivent être flexibles, auto - organisables et indépendants de la technologie de connexion. De l'autre coté, l'absence des restrictions d'accès physique dans les réseaux sans fil crée un besoin de confidentialité forte et de protection d'intégrité au niveau de la couche liaison de données. La mobilité lance de nouveaux défis à la sécurité de la communication sans fil.

Nous avons démontré notre vision des futurs réseaux sans fil, ensuite nous avons identifié les problèmes clés pour assurer la mobilité ubiquitaire sécurisée. La confiance devient le concept central pour l'élaboration des mécanismes de sécurité.

Les contributions présentées dans cette thèse visent à résoudre les problèmes suivants :

- ✓ **Extension de la région de mobilité** : la région de mobilité d'un utilisateur est limitée aux réseaux gérés par son fournisseur d'identité et ses partenaires. Dans notre modèle de confiance, la région de la mobilité de l'utilisateur peut être étendue par les relations contractuelles entre les fournisseurs d'accès réseau et par les relations de confiance entre l'utilisateur et les réseaux.
- ✓ **Sélection du réseau candidat** : l'utilisateur demande à avoir accès à ses services préférés et à les utiliser avec une bonne qualité de service. Si plusieurs réseaux sans fil sont disponibles, l'utilisateur peut choisir le réseau le plus fiable. Pour réaliser ce choix nous introduisons une méthode basée sur le classement des fournisseurs de services. Dans ce travail, nous n'utilisons pas d'information sur les réseaux.
- ✓ **Constitution dynamique des relations de confiance :** Dans notre Fast re-Authentication Protocol (FAP), nous avons éliminé la nécessité de la communication entre le réseau cible et le réseau d'origine d'utilisateur pour établir la confiance entre l'utilisateur et le réseau visité. Nous proposons un mécanisme pour la délégation de confiance basée sur les tickets d'authentification.
- ✓ **Hétérogénéité des preuves de l'identité :** le ticket d'authentification utilisé par FAP contient l'information dérivée du résultat de l'authentification précédente. Il est indépendant du type de preuve de l'identité et de la méthode d'authentification utilisée.
- ✓ **Authentification rapide pour le handover inter-domaine**: Le protocole d'authentification rapide (FAP), proposé dans cette thèse, permet de diminuer la latence d'authentification inter-domaine. Ce protocole localise le processus d'authentification, élimine la nécessité de la gestion des preuves d'identité et minimise la communication entre le domaine visité et le domaine d'origine de l'utilisateur. La méthode ne nécessite pas de stockage de données centralisé et de révélation de la topologie du réseau. L'optimisation de la signalisation liée à l'authentification combine la requête des preuves d'identité et l'actualisation de l'emplacement auprès du réseau d'origine ou du broker. Afin de minimiser le nombre de tickets générés et envoyés à chaque utilisateur nous avons proposé l'utilisation d'une table de voisins qui est maintenue par chaque fournisseur de services. Sur demande, il génère les tickets pour les réseaux placés sur la même ligne du tableau que l'emplacement actuel d'utilisateur. La méthode diminue le nombre des tickets envoyés et, par conséquent, la surcharge de trafic.

Les résultats numériques obtenus dans les expériences sur un banc d'essai et une série des simulations montrent que la méthode proposée améliore les paramètres de handover inter-domaine tels que la latence d'authentification et le coût de la signalisation.

- ✓ **Contrôle d'accès dans l'environnement ouvert :** Le mécanisme de contrôle d'accès proposé permet aux réseaux, servant plusieurs clients qui sont potentiellement malveillants, d'adapter automatiquement les politiques d'accès à la situation. Tous les paramètres de ce modèle sont définis directement par les politiques d'accès. La capacité de l'approche proposée, de réagir aux attaques a été prouvée par simulations. Le modèle de contrôle d'accès peut être implémenté dans des architectures décentralisées.

Finalement, nous avons réalisé l'intégration du protocole d'authentification rapide (FAP) introduit avant et le modèle de contrôle d'accès fondé sur la confiance.

6.2 Perspectives

Nos contributions n'ont pas inclus les schémas de paiement pour les services utilisés dans un réseau visité. En utilisant les mécanismes de facturations existants, l'utilisateur ne peut pas vérifier l'information concernant ses consommations envoyées à son fournisseur par un fournisseur visité. De plus, si un réseau offre des services pour un certain montant, les critères de paiement doivent être inclus dans le modèle de contrôle d'accès basé sur la confiance.

Nous avons proposé un mécanisme généralisé pour le contrôle d'accès. Les difficultés d'implémentation de ce mécanisme se résument surtout à la nécessité d'analyse statistique du comportement des utilisateurs avec des politiques d'accès différentes. Il est souhaitable d'élargir le contexte de ce modèle en déterminant ce que signifie « le comportement négatif ». Afin de réaliser cela, il est nécessaire d'étudier le fonctionnement des outils d'observation d'activité dans un réseau.

Dans cette thèse nous avons défini la problématique de la découverte des réseaux candidats par l'utilisateur mobile et de la préparation du handover. La poursuite possible de ce travail peut consister en la définition de mécanismes indépendants du média pour l'échange d'information sécurisée entre le réseau candidat et l'utilisateur mobile. Ce travail peut être réalisé dans le cadre du Media Independent Handover protocol.

7 LEXIQUE ANGLAIS - FRANÇAIS

Cette section a pour but de mettre en correspondance la terminologie technique en français et en anglais utilisée dans ce manuscrit.

Terme Anglais	**Terme français**
Access control	Contrôle d'accès
Accounting	Comptabilité
Authentication	Authentification
Authorization	Autorisation
Availability	Disponibilité
Candidate network	Réseau candidat (cible)
Confidentiality	Confidentialité
Credentials	Preuves de l'identité
Fast handover	Handover (transition) rapide
History	Historique
Home Network	Réseau d'origine
Integrity	Intégrité
Mobile Node	Nœud mobile
Privacy	Intimité, Protection de la vie privée
Protection	Protection
Safety	Sécurité physique (innocuité)
Seamless mobility	Mobilité sans couture
Security	Sécurité
Serving (current) network	Réseau (courant) servant
Smooth handover	Handover (transition) sans perte
Target network	Réseau destinataire (cible)
Trust	Confiance
Trusted	Fiable, digne de confiance
Trusted third party	Tiers de confiance
Trustworthiness	Fiabilité, digne de confiance

ABSTRACT

The development of wireless technologies grants a user equipped with a portable wireless device the possibility to access services any time and anywhere. Different network access technologies have been designed for different purposes. Today's digital universe is heterogeneous in various meanings of the word. Multiple IP-based services are offered for users who subscribe to multiple service providers, and have multiple roles and identities. These users are equipped with multi-interface, handheld devices with different capabilities and thus they are able to access a wide range of services over multiple access networks managed by multiple authorities. The limited scope of each access technology forces a user to gain connectivity through a verity of network technologies. For the same reasons, different technologies coexist in the same geographical areas. There is a great need for new paradigms and approaches to manage this heterogeneous universe and to deliver to users services adapted to their current terminals and access modes.

In this thesis, we study the current situation and trends in wireless technologies development. We discuss the problems related to security mechanisms specific to each technology, and in particular the possibilities for integration and interworking. Security solutions always have trust models beneath them. In the modern, dynamic, wireless world there is a strong need for trust establishment procedures. Security mechanisms to be implemented under ubiquitous mobility scenarios should be flexible and independent of operator, infrastructure and the underlying wireless technology. The key challenges to ubiquitous, secure mobility have been identified and the advantages and shortcomings of existing solutions have been analyzed.

We first study the possibility of authentication latency decreasing in a scenario where the network access authentication is decoupled from the service access authentication. An authorized user is granted network and service access as a result of a single authentication process that combines 802.1X and PANA operations.

Then we introduce the Fast re-Authentication Protocol (FAP) for inter-domain roaming, which aims to reduce the authentication delay for a mobile user in a visited administrative domain. The approach eliminates the need for communication between the target and the user's home networks for credentials verification. We develop the Fast re-Authentication Protocol by suggesting a ticket distribution scheme for inter-domain roaming. This method decreases the number of tickets sent and consequently the overhead and delay of the ticket acquisition phase of the protocol. Numerical results obtained from experiments on a test-bed and a series of simulations show that the proposed scheme enhances inter-domain handover parameters such as authentication latency and signalling cost.

To improve the access control to network resources we propose the adjustable trust model. The purpose of this work is to provide the network with the opportunity to react to user behaviour. The network is able to observe the activity of each user and to calculate corresponding trust. Clients having low trust due to illicit behaviour are not allowed to access the network. Users are motivated to gain higher trust because trusted users have access to a larger set of services with higher quality of service. Validation of the proposed trust-based access control method has been done via simulations.

Finally, we discuss how the proposed solutions can be implemented in a single framework.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

INDEX

Chapter I Introduction

I.1 BACKGROUND AND MOTIVATION

The current situation in the world of wireless communications leads to new requirements from users. Ubiquitous mobility presupposes that the user can access the Internet anytime and anywhere. Under these conditions the user needs to handover between networks managed by different authorities, and changing of access technology also can occur. Real-time applications such as voice over IP running at the mobile terminal are very sensitive to the latency of transition between the serving and the target points of attachment. Finally, mutual authentication must be performed between the mobile terminal and the target network on each handover.

Inter-domain mobility presumes that the user is involved in handover and roaming procedures. Handover is a process of changing a point of attachment when the mobile terminal moves from one cell to an adjacent cell. These cells may belong to one network operator or to several operators. Handover may occur when a session is running at the mobile terminal. Roaming is an agreement between two operators that allows subscribers of one of them to access networks managed by another one. In the case of roaming, mobile users must prove to a target operator that they are subscribed to its partner.

The significant part of the handover delay is caused by the authentication procedure. Authentication protocols in current use were designed without taking terminal mobility into consideration. When the user requests access to a visited network, the latter must communicate with the user's identity provider in order to verify the user's credentials. These communications cause delays, which are impossible to predict and to decrease. These circumstances make the authentication latency and accordingly, the handover delay unacceptable for normal running of a real-time application.

Minimization of authentication latency is an essential point for development of future mobility management solutions. Solutions for fast handover, authentication and network access control hold great interest in academic and commercial research, but there are no common approaches or proposals that cover fast inter-domain and inter-technology authentication. Effective distribution of authentication material is necessary to assure the possibility of inter-domain roaming and handover.

Mobile users can be served by its home networks (the networks where they are subscribed), roaming partners of the home network or an open public network. Open networks are subject to different attacks, but there is no guarantee that the client, which has been successfully authenticated to the network, will not manifest malicious behaviour.

In the modern wireless world the trust notion plays a significant role. Trust and reputation models are widely used in electronic commerce systems, social networks

and peer-to-peer communications. Trust models used in electronic commerce are generally based on feedbacks on one agent's behaviour provided by other agents. Trust models, implemented for network access control, are static and reflect relations between the truster and trustee only at the moment of the agreement. The example of such trust models is the case of the network partnership when one network serves subscribers of another one. The serving network trusts the client via a digital certificate signed by its partner or via the result of the login/password verification by the partner.

The access network provider is motivated to implement a kind of dynamic trust model to manage the access rights of all clients based on their past behaviour. The use of such a model permits restricting access to services for suspicious clients and provides more privileges to those clients who have demonstrated good behaviour. Well behaved clients are also motivated to participate in the trust construction because a good reputation allows them to have access to a larger set of services.

I.2 PROBLEM STATEMENT

Currently proposed solutions for fast authentication are limited to a particular access technology (like in GSM or 802.11 networks) or to a set of different access networks managed by the same operator, such as Unlicensed Mobile Access (UMA), adopted by 3GPP.

There are many limitations for security solutions deployment in the mobile environment. Firstly, the authentication and confidentiality establishment procedures must not significantly increase handover latency. To achieve this, inter-domain signalling during handover execution should be avoided, and it may be replaced by post-authentication signalling. In the event of frequent handovers among a large number of roaming partners of the user's identity provider, a large number of messages are exchanged between the mobile terminal or visited networks and the mobile user's home network, and thus the problem of traffic overhead appears.

In the mobile environment each security domain implements its own authentication mechanisms. Security solutions differ from one technology to another. Fast inter-domain authentication must be independent of the technology of the serving and candidate networks of attachment as well as of the method used for the previous authentication. Authentication solutions should not implement heavy computations due to the short battery life and limited time for handover.

For a public accessible network it is desirable to give or not give access to services for a client taking into consideration his past behaviour. This challenge may be addressed by implementation of a trust model that is able not only to deny access for malicious clients but also to adapt network and service access policies to the current risk level in the managed environment.

Thus, the framework for authentication and access control in an open and mobile environment should bring benefits for both mobile users and service providers. Mobile users should be able to move across access networks based on different technologies and managed by different authorities in a transparent manner, and the

set of services provided by a network should depend on the reputation of the user in this network. A flexible trust-based solution for network access control should decrease the ratio of service access misuse and abuse by means of access restriction for non-faire clients.

I.3 SUMMARY OF CONTRIBUTION

At present, mobility of a user is no more limited by the home operator's and its partner's networks. Public access networks deal with a great number of users. Possible trust relationships between service providers and between a user and service provider have been studied. The analysis provided of possible roaming scenarios is based on this study. An analysis of challenges and issues raised by authentication in inter-domain and inter-technology handovers is provided. The study of access control approaches has created the necessity of analyzing various trust and reputation models in order to build a more flexible and realistic access control model.

In this work the following contributions are made:

Compound authentication for wireless LAN. We propose a compound method for user authentication in a public access wireless LAN, when this requires separate authorization to access internal network services and the Internet. This approach aims to minimize the risk of attacks at network nodes conducted by unauthenticated users, provides key establishment and strong encryption between a mobile node and an access point, and decreases authentication latency. An authorized user is granted network and Internet access as a result of a single authentication process that combines 802.11i and PANA operations.

Fast re-authentication protocol. We introduce the Fast re-Authentication protocol (FAP) for inter-domain roaming, which aims to reduce authentication delay of a mobile user in a visited administrative domain. The approach eliminates the need for communication between the visited network and the subscriber's home network for credentials verification, and uses a short-lived lightweight re-authentication ticket, which does not require a revocation mechanism. FAP allows mutual generation of key material, which serves to produce session encryption keys. The proposed approach is composed of two sub-protocols: ticket distribution and fast re-authentication.

Optimized ticket distribution scheme for Fast re-authentication Protocol. Knowledge of the neighbourhood of the current network of attachment of the user may be used to reduce the number of tickets generated and sent. The scheme introduced of authentication ticket distribution reduces network load at the ticket acquisition phase and makes it possible to serve a greater number of highly mobile users. The functionality of this scheme is based on a structure called the neighbour table, which contains information about the geographical location of partner networks. We propose two modes of protocol operation: reactive, when the neighbour table is not created and the user requests authentication tickets for a chosen target network, and proactive mode, performed just after successful authentication and delivery to the user of authentication material for all partner networks reachable from the current location.

Evaluation of Fast re-Authentication Protocol. Numeric results obtained from simulations and experiments on a test-bed have shown that the proposed approach improves inter-domain handover parameters such as authentication latency and signalization cost. We have analyzed the inter-domain authentication and distribution of credentials by means of simulations and have studied the possibility of integration of the prototype developed in the IEEE 802.11 technology network.

Trust-based access control architecture. The proposed model for trust-based access control consists of the experience that the network has with each client, the recommendation of the client (e.g. certificate) and the reputation of the entity that has recommended the client. The significance of each component evolves over time; this allows the unknown user to be served owing to the recommendation from an authority trusted by the serving provider, while on the other hand a well-known user can be served even without recommendation from a trusted third party.

Adjustable observation-based trust model. We propose an adjustable observation-based trust model capable of dynamically adapting the access rights of the requested user based on past behaviour and adapting network access policies according to the currently observed risk level. In the context of this observation-based trust model we propose a concept of user history memorization and the development of the importance of behaviour records over time. We address the integration of Fast re-Authentication Protocol and Trust-based access control into a generic framework.

Finally, part of this thesis was contributed to the work of the IEEE 802.21 Security Task Group.

I.4 ORGANIZATION OF THE THESIS

This thesis focuses on the problem of fast mutual authentication and trust-based access control in an open heterogeneous environment.

Chapter II gives a generalized overview of the current situation in the world of wireless communications. We study how heterogeneous services may be delivered to heterogeneous handheld user devices over heterogeneous access technology. And we then examine issues related to ubiquitous mobility and provide classification of mobility management approaches.

In Chapter III, we inquire into the question of trust and security relationship. This chapter provides an analysis of typical security risks presented in wireless technologies and solutions to address them. We emphasize the importance of trust understanding in modern security design. We focus on the security issues brought by mobility. Access control in open mobile environments becomes increasingly challenging. Chapter III traces the history of access control mechanisms, their development and current trends.

In Chapter IV we introduce our vision of future ubiquitous, wireless networks. We concentrate on the security problems present in a heterogeneous environment. We underline the key challenges to secure ubiquitous mobility and we provide an analysis of existing solutions to overcome heterogeneity of security solutions along with related issues.

Chapter V introduces our methods for fast authentication in an inter-domain handover scenario. Firstly, we propose a compound user authentication to a visited network that combines commonly used link-layer and network-layer authentication methods. Further work on fast inter-domain authentication has resulted in the Fast re-Authentication Protocol (FAP), designed along with the scheme of roaming credentials distribution.

Chapter VI is dedicated to the trust-based access control model description. We introduce a generalized scheme for access policy enforcement based on dynamic trust evaluation. Here we also address the issue of reliable collaboration partner selection by introducing the user-oriented scheme for service provider ranging based on their trustworthiness.

Finally, Chapter VII concludes this thesis and provides an outlook on perspectives and future work.

Chapter II Current situation in the world of wireless communications

As distinct from wired, wireless technologies enable two ore more physical devices to communicate with each other without making use of a physical connection such as a cable, using instead radio frequencies to transmit data. They represent means of connectivity between an end user and a global wired communication network such as Internet or a telephone network. Wireless networks have become very popular in business and everyday life. Both individuals and organizations are finding benefits in their ease of installation and ease of use. An increasing number of Internet and service providers offer wireless access to their networks. Wireless technologies offer a wide range of capabilities oriented to different users and purposes. The benefits of wireless include connections when no others are possible, connections at lower cost in many scenarios, faster connections, networks that are much faster to install and data connections for mobile users.

Wireless technologies range from very simple personal area networks covering several feet to complex cellular systems. Wireless technologies are changing rapidly; with new features and products being continuously introduced.

II.1 SERVICE ACCESS ORGANIZATION OVER WIRELESS MOBILE ACCESS NETWORKS

II.1.1 Wireless network types

Different types of wireless access technologies may be defined based on the supported data rate, coverage area, technological factors like assured latencies, packet loss ratio and jitter, and supported quality of service classes. Various technologies are designed for different purposes. Some of them provide best support for voice transmission but poorer quality service for data exchange. Access networks may presume the existence of a back-end infrastructure and inner mobility support. The simplest and common classification of wireless access networks may be done based on their coverage area.

Wireless Personal Area Networks *(WPAN)* provide wireless interconnection of devices placed around an individual person's workspace. Typically, a wireless personal area network uses a technology that permits communication within a very short range about 10 meters. One such technology is Bluetooth, which has been used as the basis for a new standard, IEEE 802.15.

Wireless Local Area Networks (WLAN) technology provides short-range, high-speed wireless data connections between mobile data devices (such as laptops, PDAs, phones and home entertainment equipment) or between mobile data devices and nearby Access Points (AP). An Access Point may have two interfaces, wireless and wired to

provide connectivity between a wireless station and a broadband wired network. Use of WLAN as a client technology offers rapid installation, flexibility, scalability and good throughput at low cost.

Air interfaces for WLAN are defined by IEEE 802.11a/b/g standards [1]. Initially the standard did not provide mobility and quality of service support, but further development of the technology has resulted in significant extensions in the capabilities of the standard such as fast inter-AP transition, defined in IEEE 802.11r and quality of service support, defined in IEEE 802.11e.

IEEE 802.11 networks may operate in different modes. In an *Independent Service Set (ISS)* mode a wireless network operates in peer-to-peer mode, with each wireless station transmitting and receiving traffic. In *Basic Service Set (BSS)* mode the Access Point manages a cell and it is charged with traffic delivery from one station attached to this cell to another wireless station. Several BSS connected to a wired network constitute an *Extended Service Set (ESS)*. A *Distribution system (DS)* connects all the APs together, forwarding network traffic and allowing for the movement of mobile wireless stations within a much wider area.

***Wireless Metropolitan Area Networks** (WMAN)* extend the coverage area of local wireless networks and provide services, primarily Internet access, to a large number of users. New developments in wireless technologies create an opportunity to implement WMAN technology as a cheaper last mile solution. The air interfaces for WMANs are defined by the IEEE 802.16 [2] standard, also referred to as *Worldwide Interoperability for Microwave Access* (WiMax). The technology provides inexpensive broadband access and support for user nomadism and mobility. The ITU telecoms standards body has improved it as a 3G standard, part of its IMT-2000 family of protocols. For many WiMax applications, such as Broadband Internet access, no core network is needed. Most Wireless Broadband Access systems have been data-oriented until now.

IEEE 802.16e [3], based WiMax, an amendment to IEEE Standard 802.16, as modified by IEEE Standards 802.16a, 802.16-2004 and 802.16c has been developed from the beginning with multimedia services (including voice) and mobility support in mind. The access network supports 5 levels of Quality of Service permitting voice traffic to receive faster and prioritized transport. Mobile WiMax also supports hard handover and optionally supports Fast Base station Switching and Macro Diversity Handover to enable the Mobile Subscriber Station (MSS) to switch from one base station to another at vehicular speeds without interrupting the connection.

A WiMax network consists of three elements: a subscriber station (SS) or mobile subscriber station (MSS), a base station (BS) that provides radio access functionality and the Network Control and Management System. The latter includes entities supporting routing, network management, scheduling and coordination services, multimedia session management services, security and mobility services. User terminals are connected to base stations, which are connected to the network control and management system.

***Wireless wide-area networks** (WWAN)* refer to wireless high-speed communication networks covering a large geographic area. WWANs facilitate user mobility and access to home services from a remote location. Today, most wireless data communication takes place across 2G, 2.5 G or 3G cellular systems such as GSM [4], GPRS [5] or UMTS [6]. Although traditional analog networks, having been designed

for voice rather than data transfer, have some inherent problems, some 2G (second generation) and new 3G (third generation) digital cellular networks are fully integrated for data/voice transmission. 3G Systems are intended to provide global mobility with a wide range of services including telephony, paging, messaging, Internet and broadband data. The International Telecommunication Union (ITU) [7] defines the standard for third generation systems, referred to as International Mobile Telecommunications 2000 (IMT-2000) [7]. In Europe, the European Telecommunications Standards Institute (ETSI) [8] was responsible for the 3G standardization process. In 1998 the Third Generation Partnership Project (3GPP) was formed to continue the technical specification work. 3GPP2 was born out of the ITU International Mobile Telecommunications "IMT-2000" initiative, covering high speed, broadband, and Internet Protocol (IP)-based mobile systems featuring network-to-network interconnection, feature/service transparency, global roaming and seamless services independent of location.

Universal Mobile Telecommunication System (UMTS) is an example of 3G network implementation. Different services provided by UMTS have different quality of service (QoS) classes such as Conversational class (used for voice, video telephony, video gaming), Streaming class (multimedia, video on demand), Interactive class (web browsing, network gaming, database access) and Background class (e-mailing, SMS, downloading). UMTS provides support both for hard and soft handovers. A UMTS network consists of three interacting domains: Core Network, providing transit for user traffic and network management functions, Terrestrial Radio Access Network (UTRAN) and User Equipment (UE).

Mobility support in WWANs and WMANs is typically provided by internal means and requires the introduction of additional entities and protocols in WLANs. Table II.1 summarizes services provided by different types of wireless networks.

Table II.1: Wireless networks characteristics

Wireless network / Characteristics	**WWAN (UMTS example)**	**WMAN (802.16 example)**	**WLAN (802.11 example)**
Range	300 m	Optimized for 7-10 km, up to 50 km	Optimized for 100 m radius
Coverage	Outdoor	Outdoor	Indoor
Implementation	City, country	Las mile solution, city	Office, café, airport, home
Bandwidth	10 - 384 Kbps	11 - 100 Mbps	11 - 54 Mbps
Infrastructure	Yes	Yes for 802.16e	No
Services supported	Voice, IP data	Data, Voice over IP, Video, broadband applications	Limited Voice over IP, IP data
Quality of service support	Yes, differentiated QoS classes	Yes, differentiated QoS classes	No
Mobility support	Hard/Soft handover, user mobility, terminal mobility, service mobility	Hard/Soft handover, terminal mobility	No

Another classification of wireless access network technologies may be based on the underlying standard: networks may be referred to IEEE 802 and cellular or telecom operator's networks. The convergence between different network types is a strong trend. The ITU-T defines the *Next Generation Network (4G)* as follows:

"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." [7].

II.1.2 Service delivery approaches

While the first generation of the Internet was almost entirely dedicated to the transport of non-real-time data, services with strict Quality of Service requirements have now largely been adopted (e.g. Voice over IP (VoIP) and Video-conferencing). There is a large need to deliver different services such as data (e-mailing, web-browsing, ftp), voice (telephony, radio, conferencing) and video (broadcast, multicast and interactive) to a user independently of the underlying access technology. The move towards the all-IP architecture for service delivery appears to be a major trend. Two different bases for the service delivery are being developed in parallel: deployment of the new architecture called IMS and using the existent Internet infrastructure.

Various multimedia applications and services are deployed over the existing Internet infrastructure either in a client-server or in a distributed manner. Internet users enjoy not only traditional IP-based services such as e-mailing, web-browsing or file download but also services earlier associated with telecommunication operators, for example, voice communication or teleconferencing. Seamless provision of such services is made possible by introducing an overlay control layer on top of the IP based wired and wireless networks. Internet-based services deployment is a cheap solution that provides connectivity, opportunistic routing and the absence of a single point of failure. However, on the other hand, no bandwidth allocation, admission or quality of service control and end-to-end security is guaranteed. As distinct from service provisioning by an operator, a network access provider cannot differentiate services consumed by an end user.

To remain competitive with Internet, the traditional telecommunication infrastructure aims at providing the same multimedia services with better security and quality of service support. The *IP Multimedia Subsystem (IMS)* [9] is a standardized Next Generation Networking (NGN) architecture for telecom operators who want to provide mobile and fixed multimedia services. The purposes of IMS are to provide individual control and charge for individual services and to support user access to services via different access networks. The session may be established between two IMS users, two Internet users, or IMS and Internet users as well. IMS has interfaces both to cellular and Internet architectures. IMS works with any type of access networks (fixed, mobile or wireless) with packet-switch functions. Circuit-switch networks are supported through gateways. Use of IMS makes the core network independent of a particular access technology, provides application and service mobility support, allows faster deployment of new services based on a standard architecture and offers more scalability to the architecture design.

However, several issues are related to IMS deployment. Firs of all, it represents an operator-friendly solution, it acts at the expense of the consumer. Use of the 3GPP variant of SIP introduces interoperability issues while dealing with sessions run over

IETF SIP. IMS brings operators to play a central role in service distribution. The role of operators in the billing of third-party services need to be defined more exactly.

The user chooses services with regard to his preferences, such as service cost, service delivery quality or security, irrespective of the technology of the network access. This may be achieved because the user can access Internet via telecommunication operator network and can consume operator-based services via a public access network.

II.1.3 Portable Device developments

Over recent few years the design of mobile terminals has undergone revolutionary changes: the development of user equipment and enhancement of screen quality, increasing data storage empower to run applications with larger requirements. The presence of different network interfaces enables a multi-mode user terminal to choose at any given time the wireless link best suited to the user's preferences and to the application currently running at the terminal. Each wireless network standard requires the presence of a specific wireless interface at the user terminal. Accordingly, each standard has its own terminology. In cellular networks the wireless device that enables a user to establish connection is referred as to User Equipment (UE) or Mobile Station (MS), in the WiMax specification it is called Subscriber Station (SS) or Mobile Subscriber Station (MSS) and in the 802.11 specification it is referred to as a Wireless Station (STA). In the Mobile IP the mobile is referred to as a Mobile Node (MN). Considering that the ongoing 4th generation of wireless communication will be based on all-IP infrastructure, we will employ the term "mobile node" for the equipment with installed software held by a user and enabling the user to communicate with wireless networks.

The use of small portable devices introduces new challenges to services, applications and security solutions design. First of all, there is a significant difference between the capabilities of portable and fixed devices. Mobile devices typically have very limited battery life and computational power. Moreover, the expanded functionality of all mobile devices seems to have made them susceptible to many forms of attack.

II.2 MOBILITY: THE HANDOVER AND ROAMING PROBLEM

In wireless networks mobility is associated with the ability of a user to access telecommunication services from different locations and different devices. *Nomadism* (discrete terminal mobility) implies the ability of the terminal to be connected to different networks, for example, at home and in the office. Session continuity is not supported in this case. As distinct from nomadism, the term *mobility* implies continuous uninterrupted user access to a service even while changing location or device. Ubiquitous mobility is often expressed in terms of "anywhere, anytime and any device" connectivity. Mobility is also a service; its realization requires additional support from both the part of the network and the user. This support may be provided alike or not for other services.

The use of wireless devices raises mobility support requirements. Wireless does not mean mobile. A user can always move within a WiFi cell but without mobility support he cannot move seamlessly to a neighbouring cell.

Access networks that support end-host mobility, are then referred to as *Mobile Access Networks*. Network entities may be appropriated to maintain forwarding information for mobile nodes moving from one access point to another. Mobile access networks should be distinguished from another class of networks – *Mobile Networks*, which are able to change their location while retaining their internal structure unchanged.

Mobility introduces new technological challenges and increases the complexity of modern wireless systems.

II.2.1 Mobility classification

Mobility classification may be based on the item that will be replaced. Using the moving object we can distinguish terminal, session, user and service mobility.

User (Personal) mobility allows a user to be reachable on different terminals at the same logical address.

Session mobility allows a user to continue a session even when changing a terminal.

Service mobility allows a user to maintain access to his services (network speed lists, user interface configurations) while changing devices or network service providers. It should be possible for a user to update services definitions from any terminal. The user may store his preferences either locally (in flash memory or in the PDA device) or at the dedicated "home" server, that is associated with the user's address. Service mobility support is provided by telecommunication operators in Virtual Home Environment (VHE), which is a concept for Personal Service Environment (PSE) portability across network boundaries and between terminals. The concept of VHE is that users are consistently presented with the same personalized features, User Interface customization and services, in any network and any terminal (within the capabilities of the terminal and the network), wherever the user is located.

Code mobility allows software entities (codes, objects or processes) to be relocated or moved from one terminal to another during their execution.

In this work we concentrate on the terminal mobility in heterogeneous wireless networks. *Terminal mobility* allows a device to change its location while continuing to maintain all services and sessions running. The terminal changes the point of attachment, which can be an 802.11 access point, 802.16 or cellular base station. The previous and the current points of attachment may belong either to the same or to different subnets, to the same or different administrative domains and they may support the same or different access technologies.

According to the locality impact terminal mobility may be classified into two main categories:

Micromobility refers to mobility over a small area. Usually this means mobility within a single IP domain. Micromobility protocols exploit the locality of movement by confining movement related changes and signalling to the access network. Micromobility is characterized by frequent local handovers. It is provided on the link layer and does not introduce issues related to signalling. Micromobility support realization is technology-dependent; it is supported by native network elements in GSM and there are elements for mobility support proposed for 802 networks such as IAPP and 802.11r.

Macromobility represents mobility over a large area. This includes mobility support and associated address registration procedures that are needed when the mobile node moves between IP domains. Inter-access network handovers typically involve macromobility protocols. Macromobility can be provided at higher layers and it is not supported by the network technology. Mobile IP can be seen as a mean to provide macromobility.

II.2.2 Handover

The term *handover* refers to the process of call transferring from one location to another. A mobile node is involved in a handover process when it is moving out of the coverage area of one access point to the coverage area of another access point. The mobile node should be able to continue a communication session started at the initial location after reconnecting to the new attachment point. A mobile station moving out of the coverage area of one access point could re-associate to another access point, thus performing a link-layer handover.

In this thesis, points of attachment referred to are 802.11 access points, 802.16 or 3GPP base station. If both points of attachment support the same technology handover between them is *horizontal*. If there is a need to switch to a different technology to connect to the target point of attachment, *vertical* or inter-technology handover is executed. Relations between the previous and the target points of attachment determine the type of horizontal handover that will occur. Handover executed within the same administrative domain is referred to micromobility. Handover executed between access networks of the same technology managed by different authorities is referred to as *inter-domain handover*, otherwise the mobile node executes an *intra-domain handover* and is referred to as micromobility.

Cellular networks have native support for handovers between base stations belonging to the same operator, while IEEE 802 wireless networks were not initially designed to support mobility and handovers. Due to this circumstance, further classification of handover types is possible for IEEE 802 networks. If the mobile node is moving between points of attachment within the same ESS, *inter-cell* or *link-layer* handover occurs. This type of handover is the most frequent and involves only link-layer operations. When the mobile node chooses a target point of attachment in another subnet in the same administrative domain, it executes a *subnet handover*. This procedure consists of inter-cell handover operations and a new IP address acquisition procedure in a visited subnet.

Handover metrics are defined to estimate its influence on the quality of service provided in a multimedia session. *Handover latency (delay)* is the time interval between the last communication packet received (sent) at an old point of attachment and the first communication packet received (sent) at a new one. Delay can be measured in either one-way or round-trip delay.

Jitter is the variation in delay over time from point-to-point. If the delay of transmissions varies too widely, call quality is greatly degraded. The amount of jitter tolerable on the network is affected by the depth of the jitter buffer on the network equipment in the voice path. The more jitter buffer available, the more the network can reduce the effects of jitter.

Packet loss is losing packets along the data path, which degrades voice applications.

According to the values of these metrics, the handover may be fast, smooth or seamless.

Fast handover is a handover that aims primarily to minimize handover latency with no explicit interest in packet loss (real-time applications). The purpose of fast handover is to minimize the delay in handover execution.

Smooth handover aims primarily to minimize packet loss, with no explicit concern for additional delays in packet forwarding. Smooth minimizes packet loss during handover;

Seamless handover is a handover in which there is no change in service capability, security or quality. Seamless is defined as smooth fast handovers, procedures minimizing both the handover delay and the packet lost

The configuration of the mobile node and the number of network interfaces with which it is equipped determine wither hard or soft handover may be executed. In the case of "break before make" technology *hard* handover is executed. A mobile node first loses a connection to a current attachment point, and then it begins to search for a new one. *Soft* handover uses "make before break" technology, where a mobile node has simultaneous connections to both old and new attachment points and disconnects from the old attachment point after traffic is established via the new point of attachment.

II.2.2.1 Reasons for handover

Mobile terminal configuration, available services and current access technology determine different reasons for handover. Both the mobile node and the serving network may initiate handover. Independently of the entity originating the handover process, the main purpose for handover is, firstly, to avoid current session disconnection and to keep connection quality of service as high as possible. The overlapping of different network operators' coverage areas permits users to choose at any time an access network with more appropriate characteristics (bandwidth, service cost etc.).

Cellular networks are an example of networks providing mobility support. In such networks intra-domain handover are typically *network-initiated*. A user is switched to another cell if the phone is moving out of the coverage area of the serving cell, if the call capacity of the serving cell is used up or there is an overlap of two cells coverage areas. The inter-cell handover process is completely transparent to the user and does not require mobility support protocols at the user terminal.

The handover procedure may also be *mobile-initiated*. The user expects to access services with the best quality of service. Different types of wireless technologies assure the best support for particular types of applications (for example, data or voice traffic). Cost of network access differs from one operator to another and in each location there are typically several network coverage areas that overlap. The mobile node aims to always be best connected in terms of available bandwidth, suitability of the wireless technology to the application running and service cost. Inter-domain and inter-technology handovers are usually mobile-initiated.

II.2.2.2 Handover phases

To simplify handover analysis, we are splitting handover into two main steps: handover *preparation* and handover *execution*. In the first step the mobile node discovers the need for handover during the *detection* phase, and it collects information needed to perform handover during the *search* phase. If handover is network-initiated the detection phase results in a message from the network. Otherwise the mobile node needs to determine the need for handover itself. Two approaches for handover detection are possible: an *ascending* approach, in which the need for handover is detected from the lower layers, and a *descending* approach, in which the need for handover is detected based on the QoS measurement at the application layer. In the execution step the mobile node associates with the target attachment point. Finally, procedures to make the mobile node capable of receiving and sendoff sending traffic at the new location are performed. Figure II.1 illustrates basic operations performed by the mobile node during handover execution.

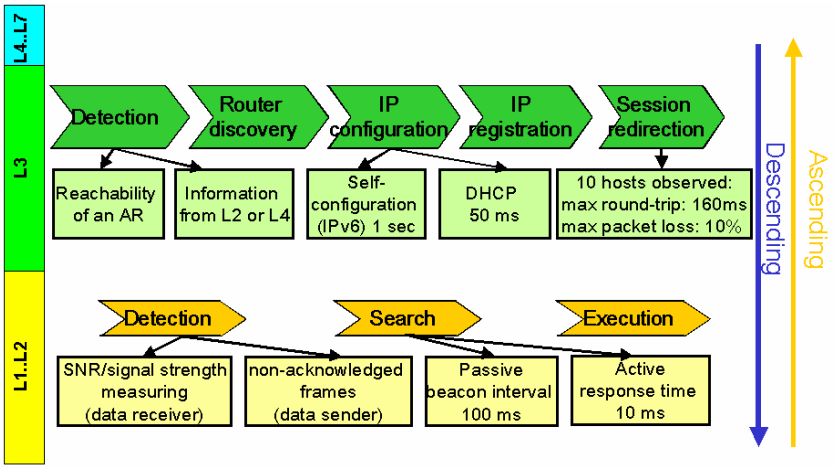


Figure II.1: Handover steps

At the physical layer, handover detection is based on permanent measuring of signal strength or/and the signal-noise ratio (SNR). At the MAC and upper layers, the handover detection mechanism depends on the role of the mobile node in communication. A mobile node can receive or send data or participate in a bi-directional session. If the mobile node acts as a data sender and does not receive acknowledgement for a sent frame, it must decide what has taken place: collision, radio signal fading or the attachment point is out of range. In the case of receiving data, the mobile node listens for the attachment point's beacons.

Methods for handover detection based on QoS degradation measurement include Mean Opinion Score (MOS) [10], Perceptual Evaluation of Speech Quality (PESQ) [11] or E-model [12]. The realization of these methods assumes the presence of a cross-layer mechanism.

After detection of need for handover, the mobile node begins to look for the most appropriate access point. The 802.11 [1] standard proposes two scanning modes for this: passive, where a mobile node listens to each channel (11 channels for USA and 13 for Europe, default beacon interval is 100 ms), or an active, faster mode, in which a station sends probe requests to each channel (response time is about 10ms) [13].

Procedures that constitute the execution phase depend on the type of handover performed. We illustrate the handover execution phase using IEEE 802.11 networks as an example. The handover execution phase is a two-step process: legacy authentication and re-association. The messages of 802.11i [14] are exchanged after the re-association process. Handover is completed when the mobile node is able to send (receive) packets from (at) its new attachment point. When *cell* or link-layer handover is performed, the mobile node is moving from one access point's coverage area to another within the same ESS. This type of handover is transparent for upper layer protocols.

Network layer protocols are involved in the operation in a case of a *subnet* or *domain* handover after link-layer handover is completed. By analogy with a link-layer handover, the need for subnet handover should be *detected*. The mobile node has executed handover to a new subnet if the serving access router is no longer reachable, a new and different access router is available, or information from a link-layer protocol (a trigger) is received. After handover detection an IP address has to be configured. This procedure may be managed either by the mobile node (IPv6 [15]) or by visited network entities (using DHCP [16]). A mobile nodes needs to determine, what service a current network s can propose and apply using the most suitable method. Network-managed address acquisition is more preferable because of the long time that self-configuration takes (the Duplicate Address Detection phase is critical for smooth operation). After receiving an IP address, a mobile node should recognize a default router to communicate with the external universe (if a network-managed approach is being employed, this stage may be skipped, since with an address a mobile node receives other information: router, DNS server address etc.). Finally, the communication session should be redirected to the new mobile node's location.

Mobility management protocols [17, 18] discussed in the following section enables the mobile node to be reachable while changing a location and a network address.

The mobile node should have the capability to determine a handover type, to decide,, what mechanisms on what OSI layers are involved in handover operation. Figure II.2 summarizes protocols that participate in an execution according to mobility type.

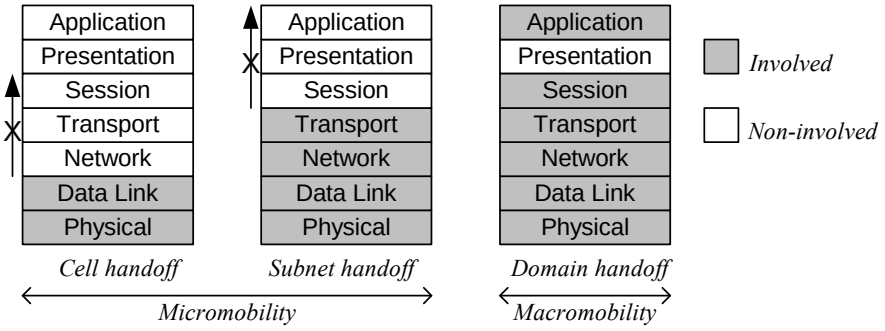


Figure II.2: Protocols of OSI levels affected by different handover types.

It is important to estimate delays introduced by the handover procedures and their impact on real-time services delivery quality, in order to design mobility management solutions in an efficient way. Handover performance at the link layer has been empirically studied in [13, 19, 20, 21]. Authors of these execution papers show that the latency is significant enough to affect the quality of service for many applications. Detection and search delay are the dominant components of the handover latency,

and a significant variation in handover parameters is observed in experiments with use of the wireless hardware from different vendors. It was observed that upstream packets are in general delayed and downstream packets are in general lost [21].

Use of two network interfaces (soft handoff), execution of handover phases in parallel [13] and improvement of each phase's performance separately [20] have been proposed and these methods permit to significantly reduce the handover latency. The reduction of search phase may be achieved by collecting information about neighbouring access point and channels during normal connectivity and use the acquired information to perform a partial active scan during a handover, as proposed in [20].

II.2.2.3 Application requirements

Requirements for handover metrics (latency, jitter and packet loss) are different for various applications running at the mobile node. *Non-real-time* applications such as web browsing or access to e-mail services do not require fast smooth operation. Handover duration for this kind of applications is limited by a timeout, after which a connection will be killed. This restriction does not play a significant role, because the waiting time to recover a TCP packet is 100 sec according to RFC 1122 [22].

Restrictions for packet loss and latency differ for various types of *real-time applications*. This kind of application is sensitive to the handover procedure that can degrade the quality of application to the point of being unacceptable for the user For *Voice over IP (VoIP)* there is no point in retransmitting lost packets. VoIP typically tolerates delays up to 150 ms, jitter less than 5 ms, and packet loss rate less than 1% according to ITU-T G.107 [23].

A *video data stream* is correlated, consisting of three types of frames for MPEG2 coding; stream decoding is impossible without key-frames (I-frames). Long handover latency increases the probability of losing the I-frame. This type of application has strict requirements for delay, packet loss rate and bandwidth.

The mobile node may use *program execution* results taken from a remote host for certain calculations. On the receiving side the data generated by a remote host may be used for other calculations. This data may be correlated and the sending part may not have a large enough buffer to store sent data. Hence there may arise a situation in which there is a need to resend dropped packets; however, they have already been deleted at the sender's side. So it is necessary to avoid a packet loss. The requirements for a handover process are to be both fast and smooth.

In the "Mobility related terminology" Request for Comments [24], roaming is defined as an operator-based term involving formal agreements between operators that allow a mobile to obtain connectivity from a foreign network. Roaming is a particular aspect of user mobility and it includes, for example, the functionality by which users can communicate their identity to the local access network. This facilitates inter-access networks agreements to be activated, and services and applications in the mobile node's home network to be made available locally to the user.

Table II.2: Application-related requirements

Services / Requirements	Multimedia video stream	Voice over IP	Data: ftp, http, telnet, SMTP	Remote program execution
delay	Strict	Strict	Tolerant	Strict
loss/error rate	Strict	Tolerant	Strict	Strict
outbound/inbound bandwidth	Asymmetric/symmetric 300Kbps/2Mbps	Symmetric 64Kbps/64Kbps	Asymmetric up to 2Mbps	Asymmetric, application dependent

II.2.2.4 Roaming

According to the definition given in RFC 1136 [25], an administrative domain is "a collection of End Systems, Intermediate Systems and sub-networks operated by a single organization or administrative authority. The components which make up the domain are assumed to interoperate with a significant degree of mutual trust among them but interoperate with other Administrative Domains in a mutually suspicious manner".

Each user must subscribe to at least one *service/identity provider*, which is responsible for the subscriber's billing. Roaming agreements between different providers create an infrastructure that allows a mobile user to gain network access via any *operator's* access network participating in such agreements.

The roaming relationship path is a set of proxies lined between a local and a home authentication server. In a large mesh of roaming partners a hierarchical forwarding model may be deployed, in which a central proxy routes the requests to their destinations. Roaming does not imply session continuity or handover between access networks.

II.3 MOBILITY MANAGEMENT

The trend in next-generation wireless systems is toward an IP-based infrastructure with the support of heterogeneous wireless access technologies. One of the challenges for these systems is the design of efficient mobility management solutions that helps to perform seamless handover. Mobility management approaches may be classified (see Figure II.3) as location management to support location registration or update and mobility management to keep a connection during changing of a point of attachment.

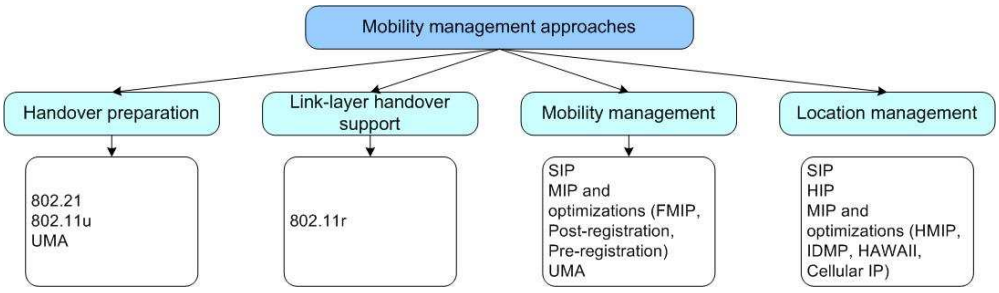


Figure II.3: Mobility management approaches classification

Mobility management approaches support scenarios where a mobile node moves

1. From one access point or a base station to another within the same wireless technology or with an access technology change;
2. Between two IP subnets within the same administrative and security domain and
3. Between access networks belonging to different administrative domains.

The second and the third scenarios may also imply changing the access technology.

II.3.1 Link-layer mobility optimization

Considerable work has been done by research and industrial institutions to optimize handover procedures in unlicensed spectrum networks. Media-specific amendments that provide various kinds of measurements, triggers and network-related information are defined. Horizontal handovers are supported in IEEE 802.11 by introducing the 802.11r amendment and 802.11f recommended practice, while 802.16e optimizes handovers between 802.16 cells.

IEEE 802.11r TG [26] is designed to support fast handover between 802.11 networks in order to allow running VoIP applications. The protocol allows a wireless client to establish a security and QoS state at a new access point before making a transition, which leads to minimal connectivity loss and application disruption**.** Under 802.11r, clients can use the current access point as a conduit to other access points. Among 802.11r proposals fast roaming using multiple concurrent associations is introduced.

IEEE 802.11f [27] provides capabilities to achieve interoperability between multi-vendor access points within a distribution system using TCP or UDP as a transport. The protocol enables access points to proactively exchange user session-related context with one another and, thus facilitate handover execution within the distribution system.

IEEE 802.16e [3] provides support for cell reselection, handover decision and initiation, and downlink transmission synchronization to the target base station. Handover decision may originate both from the mobile node and from a base station. The target base station may request information on the mobile node from its serving base station over the backbone network. As the final step of handover execution the serving base station terminates all contexts belonging to the mobile node's sessions.

Vertical handovers are executed across different types of access networks and they are more opportunistic then horizontal handovers. Many institutions are involved in the work to overcome heterogeneity. IEEE 802.11u works on interoperability issues, while handover preparation is addressed by the IEEE 802.21 working group. IETF

MIPSHOP and MOBOPTS working groups propose mobility optimization solutions based on the use on network-layer technologies. ITU proposals aim to provide user, session and application-level mobility by introducing IP Multimedia subsystem (IMS) and UMA.

IEEE 802.11u [28] is an amendment to the IEEE 802.11 standard to add features that improve interworking with external networks. The overall purpose of IEEE 802.11u is to assist the advertising and connection to remote services and it indeed intends to provide information to the mobile station about the external network prior to association. The interworking issues within 802.11u's scope are limited to Link-layer and MAC enhancements to 802.11. Specifically these issues are currently Network Selection, Emergency Call support, Authorization from Subscriber Network and Media Independent Handover Support.

IEEE 802.11e [29] introduces quality of service (QoS) to maintain IEEE 802.11 networks. The amendment enables a wireless station to learn what operational capabilities are present at each access point and to associate an access point with appropriate capabilities. With the QoS facility an 802.11 network becomes a part of a larger network providing end-to-end QoS delivery.

IEEE 802.21 [30] is not a mobility management protocol in the strict sense. The standard facilitates target network selection by the mobile node and provides handover preparation. The standard provides information to allow handing over to and from cellular, GSM, GPRS, Bluetooth and 802.11 networks using different handover mechanisms. The scope of the standard is to provide link-layer intelligence and network related information to upper layer to optimize inter-technology handovers. The 802.21 standard supports cooperative use of information held by the mobile node and within the network infrastructure. The task of a mobile node is to detect available networks while the network infrastructure is well-suited to store network-related information such as a neighbour cells list, location of mobile nodes and higher layer service availability. Both the network and the mobile user are able to make decisions about the need of handover. The handover may be conditioned by triggers and measurements supplied by the link layer at the mobile node. The 802.21 framework provides MAC layer independent messages that communicate with each media via media-specific interfaces and uses triggers defined for each technology. Once a network has been selected and the handover has been initiated, mobility management protocols take care of handover execution.

II.3.2 Network-layer mobility optimizations

The purpose of mobility management protocols is to provide ubiquitous mobility in a manner independent of the underlying technology. The primary goal of micromobility management is to ensure continuous seamless connectivity to the network within frequently occurring handovers while macromobility management aims to ensure that a mobile node can reestablish connectivity after moving to a new administrative domain.

Mobile IP is a standard that lets mobile device users whose IP address is associated with one network stay connected when moving to a network with a different IP address. Mobile IP comes in to versions: Mobile IP version 4 RFC 3344 [31] and version 6 RFC 3344 [32], Table II.3 provides a comparison of the key features of these protocols. It allows host mobility over the Internet. The protocol is based on a

network infrastructure that includes a Home Agent (HA) in a home network and a Foreign Agent (FA), which must be present in a visited network. A "Care of address" is an IP address used by the mobile node when it is attached to a Foreign network. Finally, a corresponding node is the node that communicates with the mobile node. The protocol supports mobility management by registering with a FA and location management by address binding at the HA and the corresponding node. Soft handover emulation is realized by triangular routing.

Collaboration with link-layer protocols via special messages (triggers) can improve handover parameters [33]. Pre-registration handover schemes allow a mobile node to register with a new Foreign Agent while being attached to an old Foreign Agent. Registration begins when a trigger signalling upcoming change at link-layer is received. With use of a ***Post-registration*** handover scheme registration occurs when a link-layer handover is completed. This approach is based on a network-initiated model and supposes the establishment f a bi-directional edge tunnel. *Fast Handovers for Mobile* IPv6 specification (RFC 4068 [34]) addresses network-layer handover latency reduction by introducing procedures for proactive available access router and associated access points discovery, movement detection, and proactive care-of-address acquisition. To reduce binding update latency, the protocol specifies a tunnel between the previous and the new care-of-address supported by the previous access router.

Table II.3: Comparison of IPv6 and IPv4 key features

Key Features	Mobile IPv4	Mobile IPv6
Special router as a Foreign Agent	Yes	No
Support for route optimization	In extensions	Part of the protocol
Symmetric reachability between MN and its router at current location	No	Yes
Secure operation without pre-arranged security association	No	Yes
Routing bandwidth overhead	More (encapsulation)	Less (routing header)
Decouple from Link Layer	No (ARP)	Yes (Neighbour Discovery)
Need to manage tunnel soft state	Yes	No
Dynamic home address discovery	No	Yes

Mobile IP provides high reliability (using multiple home agents) and a security layer (protection of signalling, authentication and communication with IPSec), but address binding and secure tunnel re-establishment take a long time. This approach is not scalable for frequently moving users on account of a significant handover latencies caused by movement detection, care-of address acquisition and binding update. Several modifications allow use of MIP for micro-mobility management. They reduce the number of messages sent to the home network when a mobile node changes its location in the same region. Two types of intra-domain implementations of MIP are proposed: tunnel-based (HMIP and IDMP) and routing-based (Cellular IP and HAWAII).

Hierarchical Mobile IP (HMIP) [35] and ***Intra-domain Mobility Management Protocol*** (IDMP) [36] introduce a packet redirection mechanism within a domain

using a hierarchy of mobility agents. ***Cellular IP*** (CIP) [37] and Handoff-aware wireless access Internet infrastructure (***HAWAII***) [38] use a cross-layer approach to create paths to the MN moving between APs.

Location management solution is proposed by ***Host Identity protocol (HIP)*** [39] that introduces a Host Identity (HI) namespace and a protocol layer between the internetworking and transport layers. The protocol supports readdressing and authentication services. Transport-layer associations are bound on HI, which allows host changing an IP address without modification of a transport association. The mobile node must inform the corresponding node about new address/es, and the corresponding node must verify that the mobile node is reachable at this address.

II.3.3 User and session mobility support

Personal mobility management is realized by the use of a Personal Communication Service (PCS) [40], which is the capability that provides authentication of a user and maintains user location information in the service profile, controls the completion of calls based on user-specified incoming call management contained in the service profile, provides translation between user identification and identification of the terminal currently associated with the user for the completion of calls to the use's current location, and controls the services and features available to the user based on the user's subscription and in conjunction with user-specified terminal access configurations.

Unlicensed Mobile Access (UMA) technology is defined in 3GPP. It provides access to GSM and GPRS mobile services over unlicensed spectrum such as 802.11 or 802.15. Service providers can enable subscribers to roam and handover between cellular networks and public or private unlicensed wireless networks using dual-mode handsets. The technology provides the same user identity over cellular and wireless links. UMA is a device-layer technology used by IMS to enable one mobile device to use multiple access types. IMS itself provides services delivery to subscribers in a terminal- and location-independent manner.

Session mobility management is tightly coupled to terminal mobility. After a mobile node reestablishes connectivity at a new location it should be possible to continue a multimedia session started at a previous location. ***Session Initiation Protocol (SIP)*** [18] is designed for establishing, modifying and terminating multimedia sessions. The SIP infrastructure includes a user client and a user server at a terminal, a SIP proxy, a registrar, location and redirect servers, so protocol operation does not depend on wireless and network technology. Its ability to modify sessions is used for location updating and session mobility management. For carrying out mobility management it is assumed that each visited domain has an SIP proxy. Use of hierarchical registration schemes can reduce the round-trip time of location updates.

II.4 CHAPTER SUMMARY

In this chapter we provide a classification of wireless access technologies. We analyze different types of mobility and related challenges. We give an overview of mobility management technologies in heterogeneous wireless access networks. The use of

wireless networks has increased dramatically over the last few years and their further development has led to the creation of 4th Generation ubiquitous wireless access IP-core networks, which provide providing mobility and multimedia services access anywhere, anytime and on any wireless equipment.

One of the most challenging tasks in mobility support is the reduction of handover delay and session continuity support. Much effort has been invested both by academic and industrial institutions to optimize handover execution. 3G cellular networks provide natural means of supporting terminal, personal and session mobility. They address the issue of interoperability with 802 technologies but they do not provide macromobility support.

Mobile IP provides a technology-independent solution for mobility management, but it is not suitable for real-time session continuity support because of triangular routing, which increases handover latency and a home address encapsulation, which significantly increases the overhead.

There are two essential trends in modern mobility management design: function separation between different OSI layer protocols, and tight interaction and information exchange between these layers via special triggers. Such an approach allows integration of solutions designed for particular purposes into a single flexible framework. The IEEE 802.21 working group provides a solution for media-independent handover preparation and decision; it is assumed that mobility management protocols take care of handover execution.

The combination and interoperability of link-layer, network-layer and session-layer mobility management solutions facilitates achieving good handover performance and meeting ubiquitous mobility requirements. At the same time, the use of mobility management approaches must not introduce security issues. The next chapter provides an overview of security mechanisms implemented in wireless technologies and an analysis of security issues introduced by horizontal and vertical mobility and mobility management solutions.

Chapter III Trust and security considerations

Ubiquitous use of wireless telecommunications in business and everyday life introduces new security requirements and challenges. In this chapter we provide an analysis of risks present in wireless networks. We then give an overview of security mechanisms implemented in heterogeneous wireless networks. Communications with third parties in order to verify credentials, the open nature of wireless networks, heterogeneous security solutions implemented and service consumption by users unknown in advance to a provider all bring to the necessity of a new paradigms search. Trust becomes a fundamental concept in security design. This chapter provides a brief overview of trust models, their implementation and limitations.

III.1 RISKS AND CHALLENGES IN WIRELESS ENVIRONMENT

A wide variety of network protocols coexist in a wireless world. Handheld devices are characterized by limited computational capacity and limited battery life; they may be easily stolen and can reveal sensitive information. All of security risks present in wireless network are inherited from wired networks while others are specific to wireless connectivity. All this makes the design of security solution particularly challenging.

Communication security solutions that were developed for wired networks in general are not suitable for wireless mobile communications. Wireless networks are particularly vulnerable to attacks because it is difficult to prevent physical access to them. The most significant source of risks in wireless networks is related to the communication media open to intruders. The lost of confidentiality, integrity, masquerading and denial of service are the risks typically associated with wireless networks.

The NIST handbook on computer security [41] classifies security threats in nine categories, ranging from errors and omissions to threats to personal privacy. A typical threat at the physical layer arises from the water torture attack, in which the attacker sends a series of frames in order to consume the recipient's battery. Another attack consists of jamming of a radio spectrum, thereby denying service to all parties. Finally, sensitive data may be corrupted not due to an intruder's actions but due to a transmission error. Attacks related to the physical layer are difficult to prevent and to mitigate.

Denial of Service (DoS) is a common network security problem, and it refers to an attempt to disrupt the function of a service. The disruption can range from physical destruction of network equipment to attacks that are designed to use up all of a network's bandwidth. It could even be an attempt to deny a particular person from using the service. DoS is particularly problematic in the wireless realm because of the ease of network access. DoS attacks can target both a particular station and a

network. DoS attacks are simple, but they can only achieve limited goals. Network access can provide an attacker with much greater benefits.

Man-in-the-Middle (MitM) attacks generally refer to an active attack in which an attacker interposes between two parties for nefarious purposes. In the case of wireless networks, the physical challenge is greatly reduced, and an attacker simply needs to sniff or send the right packets to perform such attacks. Wireless protocols need protections to prevent MitM attacks. Impersonating a server usually means that the attacker must set up both an intercepting access point and a back-end authentication server.

Rogue access points or base stations are unauthorized access points in a network. Network users often set them up for convenience, especially if there is no existing wireless infrastructure. Access points are cheap and easy to install in a network and are frequently set up with no or minimal security. Another potential danger is a physical intruder installing a rogue access point as a method of obtaining future access to a network.

The *eavesdropping* technique represents an attack against data or a preparation of another active attack on network entities. The attacker may intercept and analyze packets, but he may also violate packets or session integrity via message modification, dropping or insertion. The attacker can modify a packet or can inject complete packets into the data stream.

Replay attacks are also aimed at the integrity of the information on the network or the integrity of a specific session. Replay attacks are used to gain access to the network with the authorizations of the target, but the actual session or sessions that are attacked are not altered or interfered with in anyway.

Unauthorized Access is not directed at any individual user or set of users but against the network as a whole.

Communication security is often described in terms of confidentiality, integrity, authentication and non-repudiation of the transmitted data. In addition, there is a confidentiality of traffic, location, and transmitting an entity's address etc.

Any security mechanism must start with the assumption that an attacker can see everything. An intruder having a radio receiver is able to intercept messages sent on a wireless channel. Thus, security designs should include confidentiality mechanisms. An attacker can insert and modify messages exchanged between two authorized parties. To prevent this threat data origin authentication and integrity protection mechanisms should be implemented in network security design. An attacker can resent valid frames already transmitted between authorized parties. To avoid replay attacks, corresponding mechanisms should be introduced. To prevent rogue base station (access point) attacks mutual authentication should be made between two wireless devices before starting any communication. Authentication, authorization and access control mechanisms are designed to prevent attacks on network infrastructure, while confidentiality and integrity aim to prevent attacks on data.

III.2 INTRODUCING THE TRUST CONCEPT

In everyday life every interaction between persons or organizations is based on some kind of trust relationship. In the modern wireless world, the concept of trust plays a significant role because relations in the virtual world more and more reflect real life. Trust and reputation models are widely used in electronic commerce systems, social networks and peer-to-peer communications. Prior to providing an overview of various approaches to trust formalization we try to understand what is trust. In everyday life trust is an intuitively clear but very abstract concept.

The Oxford dictionary [42] defines trust as:

> *Noun: "confidence, strong belief in the goodness, strength, reliability of something or somebody".*
>
> *Verb: "have trust in – believe in the honesty and reliability of someone or something".*

Trust formalization is the subject of numerous recent academic works. Proposed models are based on different assumptions and use quite different analytic formalizations. Attempts to adopt the trust concept to the digital world have been made in the 1990s. D. Gambetta [43] in the article "Can we trust Trust?" defines trust in a more concrete manner as one agent's subjective probability with which another agent performs a particular action before this action is performed. He also introduces the dependency of trust from the context. In these terms trust may be viewed as a quantitative value. According to S. Marsh [44], trust implies a risk of some sort and it is strongly linked to confidence. He also adopts the situational nature of trust. In the work of R. Yahalom [45] different classes of trust are defined based on the different nature of tasks associated with them. While the work focuses on the trust derivation process, the dynamic nature of trust has not been considered. T. Beth, M. Borcheding and B. Klein [46] introduce a formal representation of trust relationships and a way to dynamic trust valuation. Two types of trust have been defined: *direct* trust established immediately between interacting entities and *indirect* or recommendation trust that is evaluated using reports from directly trusted entities (see Figure III.1).

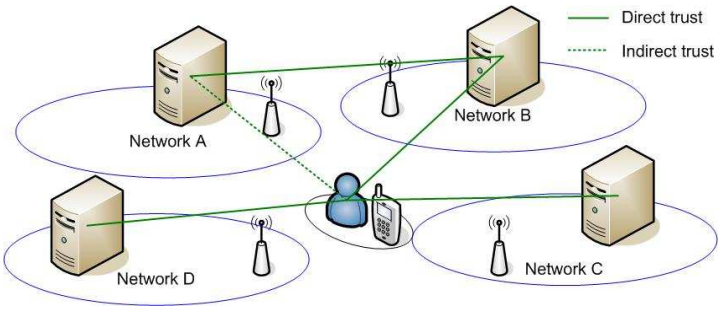


Figure III.1: Direct and indirect trust presentation

Trust relations between two entities may originate both from the reputation acquired over past interactions or derived from trusted parties' recommendations, and from contractual relationships. Contractual relationships are established with a whole domain, for example, between a user and a service provider, while trust relationships derived from this contract are established between the user equipment and particular

entities in the network managed by the service provider such as authentication or application servers and access points. Trust relationships usually exist in the form of security associations.

Trust relationships established between contract partners are mostly *static* and reflect trust that exists at the moment of entering into a contract. A roaming user soliciting services from a partner of his subscription provider *dynamically* establishes trust relationships with the latter. Thus, trust in a particular situation is created based on multi-layer relationships and is formed using a combination of different trust values. In the example we gave, the trust between the user and an authentication server in a visited network is based on a combination of trust between the user and its home provider and between the home and the visited service providers.

However, when trust is constructed using different sources, important trust characteristics [47] must be considered:

- ✓ Trust is *not symmetric:* if "agent A trusts another agent B", it does not mean that "agent B trusts agent A".
- ✓ Trust is *not distributive*: if "agent A trusts (agent B and agent C)", the statement "agent A trusts agent B and agent A trusts agent C" is not true.
- ✓ Trust is not transitive: if "agent A trusts agent B and agent B trusts agent C", it does not follow that "agent A trusts agent C".

There is a strong relationship between trust and security. A certain level of confidence and reliability assurance should be provided for any interaction between electronic devices.

Security is defined as follows:

Noun: The state of being or feeling secure; the safety of a state or organization against criminal activity such as terrorism or espionage; a thing deposited or pledged as a guarantee of the fulfillment of an undertaking or the repayment of a loan, to be forfeited in case of default; a certificate attesting credit, the ownership of stocks or bonds, etc.

Generally, security is implemented to ensure the *safety* of a system, in other words, to make a system secure against threats. Safe means

***Adjective**: protected from danger or risk; affording security or protection.*

Safety is a closely related term denoting

Something designed to prevent injury or damage.

Since IT security is expensive, it is implemented only when the potential loss in a case of non-implementation will be greater than the cost of security implementation. Security mechanisms are designed to protect infrastructure and information. Usually security objectives formulation is based on the analysis of trust to all possible entities with that a system may potentially interact. Figure III.2 shows relationship between trust and security. Each computer system manages a set of assets accessible for different agents. To decide which security solution must to be implemented to protect each asset, risks associated with each asset and trust for each type of agents should be estimated. If no risks are discovered, agents are considered trusted and no security solution should be implemented. Otherwise, security objectives and security mechanisms are defined for each asset based on trust for agents. The final security

solutions are determined by security mechanisms defined at the previous step and by the cost of their implementation.

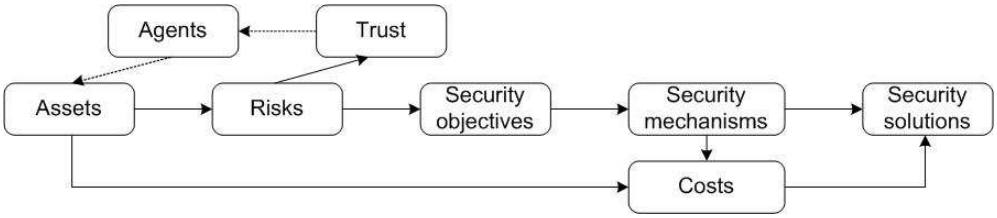


Figure III.2: Relation between trust and security

Initial insecurity of internet TCP/IP protocols is explained by the fact that these protocols where designed to be used in trusted environments, where no risks were assumed and communicating entities trusted each other. When designing a wired local office network we trust that it is difficult for an untrusted entity to intercept information exchanged between computers inside this network. That is why we will not implement confidentiality mechanisms at the link layer. In a wireless network the previous assumption is no longer valid, and therefore confidentiality mechanisms should be implemented at the link layer.

Security solutions always have trust models beneath. In such a model trust is associated with a particular action: an authentication server is trusted to keep users secrets, access points are trusted to generate session keys and a user is trusted to access certain resources based on the identity presented to the authentication server. A specific security mechanism should provide means to enable communicating entities to trust cryptographic keys.

III.3 OVERVIEW OF SECURITY MECHANISMS IMPLEMENTED IN WIRELESS NETWORKS

To design security solutions, the trust relationships between the entities involved into communications must be identified. A user is typically subscribed to at least one network access provider and they trust each other via some shared secret contained in a handheld device, for example, IMSI in a SIM card for 2G/3G subscribers or in an X.509 certificate or login/password pair at the subscriber's side and in an HLR/Authentication Server at the provider's side. Moreover, a user may choose to trust a network to transport his sensitive data only if this network provides confidentiality and integrity mechanisms.

Due to the physical properties of wireless networks, attackers will always have access to the wireless component of the network. To prevent attacker access to the wired component of a network solutions for data confidentiality as well access control mechanisms must be developed.

III.3.1 Security in WLAN: the IEEE 802.11 example

The openness of wireless communications to eavesdropping leads to the need for development of data confidentiality, data origin authentication and integrity protection mechanisms. As the first attempt to realize such a security mechanism the

Wired Equivalent Privacy (WEP) was defined in the original IEEE 802.11 standard [1]. The same algorithms have been widely used with 40-bit and 104-bit keys, referred as to WEP-40 and WEP-104. WEP uses an RC4 algorithm to encrypt the data passed over a network. The insecurity of WEP, which can be seen in easy message decryption, short and static pre-shared secrets (usually entered by the user) and initialization vectors collisions was proven by the attack simulation in [48,49,50] and recognized by the IEEE 802.11 working group.

To be able to communicate with other network entities, a mobile node must first associate with an access point. A network does not trust any mobile node and a mobile node does not trust a network prior to association completion.

The original IEEE 802.11 standard defines two types of authentication: Open System authentication and Shared Key authentication. Authentication is used between a station and an access point in BSS operation mode or between two stations in IBSS operation mode. Open System authentication is a null authentication algorithm and it involving a two-step exchange sequence. Thus, any client can authenticate itself to an access point. The Shared Key authentication uses the WEP key to authenticate a client. Both the plain-text challenge and encrypted result are transmitted over the air. In addition, the authentication is unilateral, so the client is not able to authenticate an access point.

These security mechanisms are referred as to pre-RSNA (Robust Secure Network Association). They failed to meet their security goals and therefore became obsolete. The use of wireless networks in the public sphere requires a higher level of security based on strict network access control and data communication encryption. These requirements were satisfied in the IEEE 802.1X standard [51] which implements the media-independent Extensible Authentication Protocol (EAP) [52]. The authentication and access control process involves three entities: a wireless user's device called supplicant, an access point that plays a role of authenticator and a back-end Authentication Server (AS).

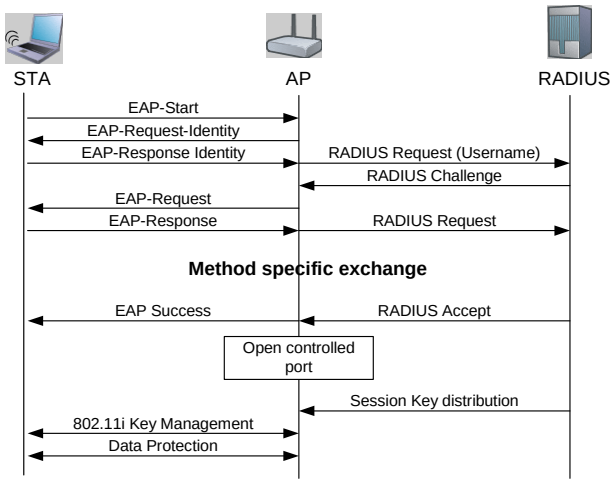


Figure III.3: 802.1X port-based authentication call flow

The mobile node trusts an authentication server via some shared secret and it trusts an access point via the authentication result. It is assumed that trust is established between the authentication server and all access points in a network and these trust relationships assume the presence of a secure communication channel between them.

An access point communicates with the supplicant over EAP and it usually transmits supplicant's messages to the authentication server over the RADIUS [53] protocol. The authentication call flow is shown in Figure III.3.

For key management, the IEEE 802.11i security amendment [14] specifies the key generation and distribution scheme. The key management phase includes two handshake types between the supplicant and the access point: 4-Way Handshake (see Figure III.4) to derive keys for unicast traffic protection, and an optional 2-way Group Key Handshake that provides means to protect multicast and broadcast traffic. The key agreement phase follows successful authentication. All keys defined in 802.11i have a limited lifetime and they are organized into a key hierarchy. At the top of the hierarchy the 256-bit Primary Master Key (PMK) is defined. The PMK is obtained either from a static Pre-shared Secret Key (PSK) stored at the access point and the supplicant or from a Master Session Key (MSK) derived as a result of a successful EAP authentication.

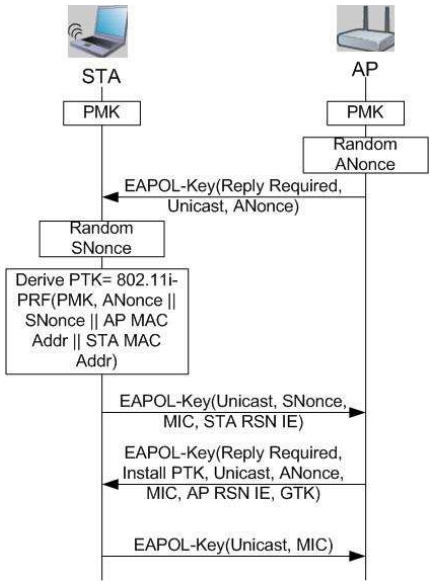


Figure III.4: 802.11i 4-way handshake

The use of mutual authentication is important in a wireless network. This will guard against many security issues, such as man-in-the-middle attacks. After strong mutual authentication both communicating entities trust each other and encryption and integrity protection keys are derived as a result of this authentication.

III.3.2 Security in WMAN: the IEEE 802.16 example

The WiMax security model is based on the assumption that there are trust relationships and security associations established between all equipment controlled by an operator.

Security services provided by the 802.16e [3] standard include mutual authentication between the subscriber station and the network, key negotiation, traffic encryption and integrity protection. Cryptographic transforms are applied across connections between a subscriber station and a base station.

802.16e security has two components: an encapsulation protocol for securing packet data and a key management protocol (PKM) providing secure distribution of key material from the base station to the subscriber station. The PKM protocol exists in

two versions: PKM Version 1 and PKM Version 2. The latter has more enhanced features such as new key hierarchy. Introducing -based device authentication with a subscriber module to the key management protocol enhance the basic security mechanisms. The key management protocol allows both mutual and unilateral authentication. Authentication is performed with

1. EAP [52];
2. X.509 [54] certificate together with an RSA public key encryption algorithm;
3. A sequence started with RSA and followed with EAP authentication algorithm.
4. The standard allows operator-specific credentials to be used in PKM besides of X.509 certificate.

The PKM establishes a shared secret referred to as an Authorization Key (AK) between the subscriber station and the base station. This key serves the secure exchange of a Temporary Encryption Key (TEK) that is the basis for 128-bit Key Encryption Key (KEK), 160-bit HMAK Key for Uplink and 160-bit HMAC Key for Downlink derivation.

The PKM RSA authentication protocol uses X.509 digital certificates and the RSA public-key encryption algorithm that binds a client's public key to the MAC address of the mobile node. The product of EAP authentication is a Master Session Key (MSK), which is known by the subscriber station, the authenticator and the authentication server. After the successful initial authentication, SS and BS shall perform re-authentication by PMK/PMK2 lifetime. In performing re-authentication, SS and BS perform double EAP just like initial authentication. Otherwise, SS and BS can perform EAP once.

Security associations are used to keep key material that is used to protect unicast communications. Each security association has a unique identifier SAID within a base station.

III.3.3 Security in WWAN: the UMTS example

The security functions of UMTS [55] are based on what was implemented in GSM [56]. Some of the security functions have been added and some existing functions have been improved. Encryption algorithms are stronger and they are included in the base station to radio network controller interface, the application of authentication algorithms is stricter and subscriber confidentially is tighter. Security elements inherited from the GSM standard include subscriber authentication, subscriber identity confidentiality, removable subscriber identity module (SIM) and Radio interface encryption. 3GPP Technical Specification 33 105 [57] defines authentication and key agreement algorithm requirements.

The UMTS Authentication and Key Agreement protocol (AKA) was designed to enforce network access security and provide compatibility and interoperability with GSM/GPRS networks in order to ensure global roaming for UMTS subscribers.

In the GSM security model a mobile node implicitly trusts a base station to which it associates. Such an assumption entails the risk of rogue base station attack. Security against the use of false base stations was improved by introducing mutual authentication between the subscriber module and the base station. In GSM

implementation user traffic was encrypted only on the air interface between the subscriber module and the base station, while in the UMTS standard traffic between a base station and the Radio Network Controller is also encrypted. Traffic between nodes in the core network is not encrypted.

UMTS specification has five secure features groups:

1. Network access security: no access is provided before mutual authentication; user confidentiality support means identity, location, signalling and data confidentiality.
2. Network domain security is operator-dependent and it is not defined in the present specification.
3. User domain security provides secure access to mobile stations and secure communication between SIM and handheld devices;
4. Application domain security features enables a user and applications to exchange messages securely.
5. Visibility and configurability of security means that the user should be informed about the presence of link encryption and the security level particularly if the user roams and intends to keep his security preferences.

Access to the USIM is prohibited until the user has been authenticated for it. A link between the user terminal and the USIM is also protected by a secret stored in a secure manner in the USIM and the Terminal Equipment. Mutual authentication should occur at each connection set-up between the user and the network. Two authentication mechanisms are defined in UMTS: between the user's home environment and the serving network and the local authentication between the user and the serving network. The authentication procedure establishes a secret cipher key and integrity key between the user and the serving network.

For confidentiality of data on the network connection link, cipher algorithm negotiation, cipher key agreement and confidentiality of user and signalling data are provided. To provide data integrity the user and the network are able to agree on the integrity algorithm, integrity key and data integrity and origin authentication of signalling data.

III.3.4 Security mechanisms in public access networks

With development of the IEEE 802.11 standard wireless networks, open public networks also called hotspots have become very popular. These wireless networks deployed in airports, railway stations, public parks, cafés and pubs make access to the wireless component open to every user and they implement authentication, authorization and access control at the network access server, a gateway device performing filtering of IP traffic. The security mechanisms are implemented either on the network or at the application layer. After association with an access point a user is either automatically redirected to a service provider's portal page to perform web-authentication [58] or a user terminal starts IP-layer authentication straight after IP address acquisition using, for example, Protocol for Carrying Authentication and Network Access (PANA) [59].

If access to network services relies on web-authentication, a user enters his credentials on a portal page run at a gateway, and the gateway uses the RADIUS protocol to

communicate with a user's service provider. This scheme is not considered to be a secure user authentication method (unless used in conjunction with some external secure system such as SSL/TLS), since the user name and password are passed over the network as clear-text in the Basic authentication. It is also preferable to run Digest authentication in an Https connection.

The Protocol for Carrying Authentication for Network Access (PANA) provides network-layer transport for Extensible Authentication Protocol (EAP) to enable network access authentication between clients and access networks. It is a protocol for a mobile node's authentication for the nearest access router and does not depend on a link-layer carrier.

Upper-layer authentication protocols do not provide link-layer security mechanisms such as encryption or integrity protection. Furthermore, they allow access to internal network entities for any user.

III.4 SECURITY VERSUS MOBILITY

In the traditional digital world the main security requirements such as confidentiality, integrity, availability and non-repudiation were addressed by installing firewalls at network borders and security level monitoring. Fixed communication may be seen as a static environment having more or less defined threats, borders, lists of communicating entities and identities and static, mostly direct trust relationships between them. The arrival of mobile communications brings a dynamic aspect into the digital environment and extends security-related requirements. Not only confidentiality of data but also confidentiality of location, traffic and identity should be addressed. The use of cryptography becomes vitally necessary but difficult to implement because of the limited processing capabilities of mobile devices.

III.4.1 Security challenges introduced by mobility

The need to maintain a security state while moving among networks introduces new threats to security solutions, besides that, differences between capabilities of portable and fixed devices are not taken into consideration for authentication and security related operations design.

Networks access providers and service providers may trust each other on a contractual basis, forming either bilateral roaming agreements or multilateral federations. In such a way a user may connect to a network managed by an authority with which he has not established direct trust (see Figure III.5). Since trust relations are not transitive, a special trust infrastructure and mechanisms are required to establish indirect trust between a roaming user and a visited network.

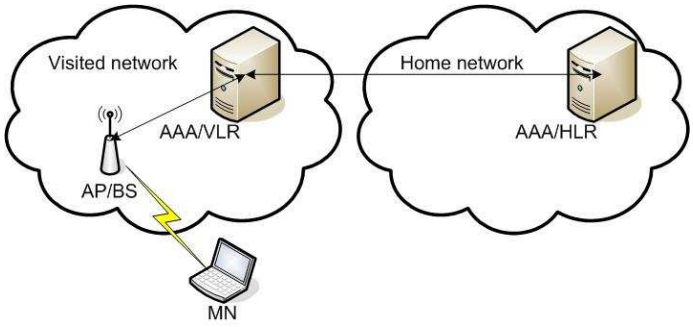


Figure III.5: Authentication of a mobile node in a visited domain

Global mobility may be viewed as a combination of a handover and a roaming problem. Roaming introduces issues related to user credentials management, indirect trust establishment and mutual authentication between previously unknown and untrusted entities. Inter-domain handover execution involves numerous entities that bring all threats associated with them into a process. Many issues are raised by handover preparation and initiation phases. As handover may be both mobile-initiated and network initiated, the risk of false handover should be addressed. When choosing a new network of attachment, a mobile node should be able to learn its capabilities and the security level provided. Independently of the application running, a mobile node passes through typical states. Table III.1 summarizes threats associated with each handover phase and possible remedies.

Table III.1: Potentially vulnerable states and solutions to secure them

State	Threat	Solution
Handover detection	False handover	Use handover-related information coming from trusted sources
Search for a candidate network	No guarantee on capabilities advertised	Work in progress
Association with a new access point	Access point impersonation	Mutual authentication with an access point
Router discovery, IP address changing	Access router impersonation	Mutual authentication with an access router
Registration with a home server	Credentials, location interception and modification	Traffic encryption and integrity protection
Session redirection	DoS on the correspondent node, session redirection, session hijacking	Authenticating and ciphering of redirection massages
Session establishment	Credentials and keys interception, entity impersonation	Encryption of signalling information, strong authentication
Session running	Eavesdropping	Encryption of communication

Secure channel binding must be provided in order to assure that end-points of a secure channel are the same as were authenticated by each other. This requirement becomes very important since a masquerade attack is easily realized in a mobile scenario.

For its part, a visited network should identify and authenticate a user that asks the network to grant Internet access; dynamically build trust relations with other administrative domains in order to authenticate a visitor; distribute/negotiate encryption keys with users; correctly account user's activity and recognize malicious behaviour in a visitor.

III.4.2 Identity management problems

To build a secure computer network it is critical to know what entities can access what resources, from where and when. Identity management represents a complex process of entities recognizing, definition of privileges and restriction for them and access granting according to these privileges. Identity management system thus involves authentication, policy enforcement, authorization and access control. In the today's communication universe interaction with multiple identity and service providers becomes an important part of Identity management. All parties involved into identity management must have specific trust relationships. Entities verifying credentials often do not have direct means to estimate trustworthiness of the corresponding entity. Systems usually rely on authentication to build trust. Trust models must take into consideration the conditions in which a party can trust others for security and privacy.

E. Damiani et al in [60] present an overview of general identity management issues such as lifecycle management, representation formats, cross-domain communications, architectural patterns such as devices and media, administration, anonymity support and dependability and trust management. Mobility introduces new trust issues to identity management. Authors of [61,62,63] propose an analysis of trust-related identity management issues. The simplest case is presented by an isolated (pairwise) identity management when the same entity is an identity provided and a service provider. In a mobile environment a user is often served by an entity that is not his identity provider and trust establishment between a user and a service provider becomes a necessity. The authors distinguish three types of indirect trust establishment in Identity management:

- ✓ Federated (community): there must be trust relationships between the identity and the service provider to allow a user to be served by the service provider. Mapping between different identifiers owned by the same user is also required;
- ✓ Centralized (brokered) where a single identity provider is used by different services. There must be service provider's trust in a credentials provider.
- ✓ Personal: all identifiers are held in a Personal Trusted Device, which is trusted and controlled by a user.

A critical issue for clients is a service provider's identity management. This issue is solved by the use of mutual authentication between the user and the service provider in order to verify is the service provider really has the expected identity. The modern ubiquitous environment represents a mixture of all cited types of identity management approaches with issues inherited from each approach.

Indirect trust establishment is based on an agreement to a common set of policies and procedures set between service and identity providers. In this scenario an identity provider delegates trust to a user, policies and responsibility to a service provider. In this scenario issues related to policy mismatch and asymmetric trust relationships appear. Trust establishment always has a cost.

To be globally reachable, a user should be uniquely identified over the Internet. Each user has multiple roles and multiple identities depending on the type of an access

network that he is connected to, the relations with this network and the user (home/visited authority). Maintaining multiple identities as separated names poses huge management problems. Numerous identities owned by a user complicate the task of activity tracking and appropriate trust evaluation. The trust value is bound to a particular identity, but there is no way to bound it to a concrete user. A user may utilize different identities to access the same service.

Moreover, inter-domain roaming introduces a requirement for end-user identity hiding from external entities [64].Ubiquitous networks require new approaches to user's identification. Traditionally a user's identification is realized at the physical (MAC), the network (IP address) or the transport (socket) OSI layer. At present, such identifiers have no further practical meaning. Since a handheld device is equipped with multiple wireless interfaces the physical device address cannot serve as a unique identifier. A mobile node acquires a new IP address in each visited network, thus an IP address no longer identifies a user.

In a cellular network a telephone number uniquely identifies each user. On the Internet user identifiers are service dependent, and may be for example an e-mail address or a Skype login. To overcome identity heterogeneity and translation problems, various solutions have been proposed. Host Identity protocol [39] experimental specification introduces a new type of identity that is independent of neither the physical nor the IP address of a mobile node. Implementation of this type of identity requires some changes in behaviour of transport protocols and it does not meet privacy-related requirements.

To address this challenge the IMS specification proposes a technology independent solution that is based on the use of two identities: IP Multimedia Private Identity (IMPI) used in the home network, and IP Multimedia Public Identity (IMPU) used outside of the home network. Both these identities are formed using URI (Uniform Resource Identifier) syntax [65].

III.4.3 Trust establishment during handover

When a mobile node is connected to a network, it trusts it to deliver traffic and provide some services. The mobile node trusts an authentication server directly via a contract if it is the user's home server or indirectly if it is a visited authentication server. Indirect trust relations are based on some information exchanged between the user's home and the visited authentication servers. The mobile node trusts an access point to which it is associated and does not trust others. If the mobile node handovers to a candidate access point belonging to the same administrative domain as the previous one, the trust relationships between the mobile node and the access point are established based on proof of knowledge of a common secret delivered by the authentication server trusted by both parties.

In case of inter-domain handover the mobile node trusts neither the access point nor the authentication server located in this network. Trust in a new domain is constructed from trust relationships established before handover using a special mechanism. Successful mutual authentication between the mobile node and the visited authentication server is possible only if there are trust relationships and security associations established between the visited and the mobile node's home domains. Usually, authentication servers communicate with one another using an authentication protocol such as RADIUS or Diameter. First, trust relationships are

established between the mobile node and the visited authentication server via the result of mutual authentication, and than the mobile node starts to trust the access point via keys derived from the authentication result (Figure III.6).

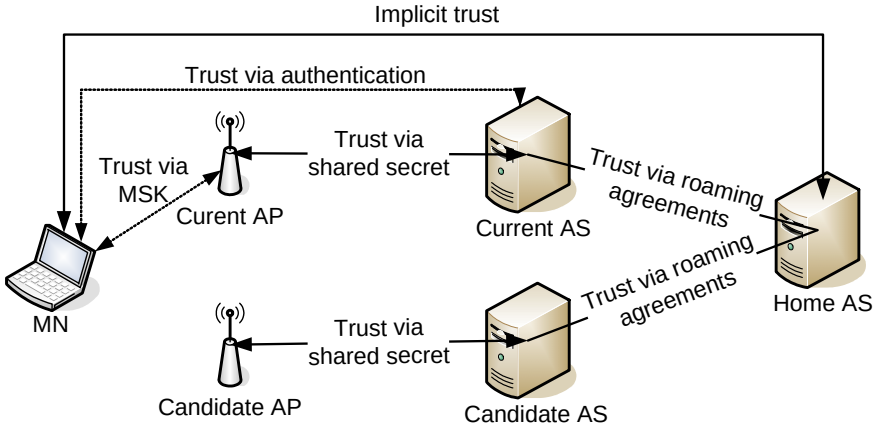


Figure III.6: Trust relationships in 802.11 in a handover scenario

To be able to communicate with other entities the mobile node must acquire an IP address and establish trust with an access router in the visited network. Typically, if a port-based authentication is implemented, the authenticated mobile node becomes trusted by internal network entities.

III.4.4 Security impact on mobility performance

For each handover scenario, two closely related aspects are very important, latency and a security level. Mutual authentication between a mobile node and a visited network presents a key issue due to its significant contribution to the total handover delay. Table III.2 summarizes the impact of each phase on the total handover latency. It is assumed that all steps except authentication are optimized with relation to latency minimization.

Table III.2: Latency of each handover phase in 802.11 network [13, 19, 20, 21]

Phase	Time, ms	%
Detection	3	0.18
Searching	40	2.35
Execution	2	0.12
IP address configuration	200	11.73
Session redirection	360	21.10
EAP-TLS authentication	1100	64.52

EAP-based authentication is widely adopted in IEEE 802 networks. The main advantages of the protocol are technology independence and flexibility. At the same time, EAP methods were designed without taking into consideration inter-domain roaming and application session continuity support. To allow a normal execution of a Voice over IP application on the mobile node, the maximum handover duration must not be more then 150 ms according to [10], while our experiments and [19,66] have shown a latency of about 1000 ms just for 802.1X authentication using the EAP-TLS method. Even if the authentication server is situated in the same network as the authenticator, the full EAP exchange time exceeds the admissible limit. Another

problem is the situation where the authentication server capable to verify user credentials is situated several hops ahead of the current authenticator.

There are two fundamental approaches for authentication between the user terminal and the visited network. The first one requires communication between the visited and the home networks during the authentication process to verify user's credentials. These communications cause delays that are difficult to predict and to shorten. The second group of approaches is based on public key cryptography. Using X.509 certificates [54] may eliminate the necessity of inter-domain communication during the authentication process, but in this case authentication is only possible if both the mobile node and the visited network recognize each other's certification authorities. Disadvantages of the method are the heavy computational cost of asymmetric cryptography and the need for a certificate revocation mechanism.

III.5 ACCESS CONTROL DEVELOPMENTS

To protect confidentiality, integrity and availability of the system the access control mechanism must be defined. *Access control* is the process of limiting access to the resources of a system only to authorized programs, processes, users or other systems. Access control is based on the definition of permissions called access policies. A security policy is a list of specific constraints aiming to protect a special system.

The earliest security policy model was proposed by D. Bell and L.G. La Padula in 1973 [67]. It is also known as Multi-Level Security. The basic properties of this model are that information cannot be read downwards and write upwards. This model addresses confidentiality issues that are constraints on who can read the message, while integrity represents a constraint on who can modify the message. The model that addresses integrity issues and ignores confidentiality was proposed by K. Biba [68]. This model is often referred as "Bell La Padula upside down".

When systems are designed to enforce security policy independently of user action they are described as having *mandatory access* control as opposed to systems with *discretionary access control* where users can make their own decisions about access rights to their files.

Access control systems typically authenticate principles and then solicit access to resources. Access control rules are usually expressed with the help of access policies. In the simplest case access policies are expressed by an access matrix [69], with columns for resources and rows for users. In such systems all users are known to a system in advance. Access control matrices do not provide a scalable solution. It might not only impose performance problems but also increases the probability of an administration mistake. To store access policies in more compact and efficient way several solutions have been proposed.

Access Control Lists (ACL) are defined together with the corresponding resource. ACLs are suitable to situations where access policies are set in a centralized manner, but they are less suited for the situation where the system interacts with a large amount of users that is continuously changing. ACLs are simple to implement, but they are not efficient for doing security checks in runtime.

Capabilities (also referred as to tickets or certificates) facilitate security monitoring in runtime and allow delegation of capabilities. Usually capabilities take the form of public key certificates. A public key certificate is a credential signed by some authority, which declares that a certificate holder is a certain person, a member of some group or the holder of a privilege.

The approaches described of access policies definition are used in different types of access control mechanisms that may be classified as follows.

Identity-based access control (IBAC) [70] defines what actions each subject (user, program, or other system) may perform on each object (file, resource etc.). As this type of access control is based on the user's identity, authentication must precede the access control. Authorization determines to what resources the authenticated user has permissions to access and the access control mechanism opens up access to these resources. The IBAC provides direct control of each subject access to each object. Systems that offer services only to subscribers and known users such as cellular operators or enterprise networks maintain a profile for each subscriber. Usually identity profiles group users with the same permissions in order to simplify administration tasks. Permissions may also depend on where the user is connected to the network: from the office, from home or from a public place. Such information is kept in a connection profile. The Identity profile and the connection profile determine together the access policy for a specific user at a specific moment.

The IBAC does not offer a scalable solution for systems where the access policies change over the system lifecycle.

Role-based Access Control (RBAC) introduced by Sandhu et al in 1996 [71] and standardized by NIST in 2001 [72] is getting the most attention at present [73,74]. This model represents more generalized version of mandatory access control where permissions are given based not on user names but on the functions that they perform or, in other words, roles. Permissions associated with a role are more stable then permissions associated with a particular user. In a role-based access control a "group" refers to a list of users and a "role" refers to a set of access permissions that one or more users can have for a defined period of time using a defined procedure. In role-based access control systems notions of groups and roles are combined. Users establish sessions during which they may activate a subset of roles they belong to. Each session maps one user to possibly many roles. Authorization management is significantly simplified in comparison with IBAC.

In IBAC and RBAC actions are elementary such as read, write and execute. Modern applications make demands for more complex actions, such as a single sign-on to a system.

Wireless networks use a shared medium and therefore, access control methods used in wired networks are not suitable for them. The additional challenge in wireless networks is that the user can appear anywhere in the network. Network security that relies on physical constraints is no longer effective. Policies for controlled networks access becomes very important with spread of wireless networks.

Access control mechanisms operate at a number of levels in a system, from the physical to the application layer. These systems are vulnerable to environmental changes that make assumptions used in their design invalid. The dynamic nature of ubiquitous networking is the most challenging aspect of access control design. Using

RBAC simplifies policy management; however several issues appear when the access control needs to be implemented in an open environment where the number and identities of users are not known in advance. The way in which the role has to be assigned to an unknown user is not clearly defined yet. The trust notion may help to formalize and automate the procedure of role association with a particular user. The protection properties should be realized by mechanisms that are simple enough to verify and that changes rarely.

III.6 TRUST MODELS OVERVIEW

A variety of trust models are proposed for P2P, ad-hoc networks and electronic commerce systems. It is necessary to determine whether they are suitable to be implemented in ubiquitous wireless networks that serve mobile users. Trust models designed for ad-hoc networks are oriented to neighbour discovery and safe routing providing. The question of roaming agreements management and access policies localisation is not yet well explored in literature.

Existing models for trust and reputation may be classified according to their purpose as trust models for domain policy mapping, trust models for choosing a reliable partner for collaboration and trust models for access control. The principle difference consists in the sources of trust used such as recommendations, feedbacks from other entities or personal observation.

III.6.1 Choosing a reliable partner for collaboration

Such models addresses trust and reputation management in P2P networks, overlay networks, Grid services, systems for electronic commerce (for example, eBay) or situations where a user is able to choose a service provider. As these trust models are designed for distributed environment, their distinguishing features are mechanisms for trust propagation, indirect trust calculation and trust delegation. In this case not only previous interactions but also the expected utility of an action, its costs and benefits, are taken into consideration for trust computation.

The Beta reputation system [75] is proposed to calculate reputation value for a peer in a decentralized system. This trust model is based on using the beta probability density function designed for use in electronic commerce. All events (feedbacks) are considered to have only two outcomes: positive and negative.

M.Srivasta, Li Xiong and Ling Liu [76] developed a reputation system called TrustGuard. This model is designed for decentralized overlay networks and makes use of three groups of information: current reports, history of the node behaviour and sudden changes in the trust value of a node in the very recent past.

N.Griffiths [77] developed an approach that combines experience-based and recommendation trust. This formal model introduces multidimensional trust. The value of each trust dimension is updated recurrently using weighting specified for successful and failed transactions. In addition to a trust value each agent has a confidence level that increases following each successful transaction.

The work of S.Park et al. [78] proposes a model that uses a trust and penalty vector to calculate service provider reputation in a web service management environment.

The purpose of the SECURE project [79] (Secure Environments for Collaboration among Ubiquitous Roaming Entities) is the development of a computational model of trust that will provide the formal basis for reasoning about trust and for the deployment of verifiable security policies. This trust model mostly addresses peer-to-peer and distributed structures.

The work of N. Dimmock et al. [80] provides an analysis of risk models for trust-based access control. They consider trust as an abstract concept and propose to use risk as a more quantified concept. They also define access control policies as a function of observed utility. In [81] decisions to serve clients are made on the basis of trust and risk instead of using credentials alone. In addition to the previous experience the outcome cost for the action influences decision-making.

A group of works proposes trust models for file sharing systems. Normally, negative experiences have more effect on trust than positive experiences. [82] uses direct trust formalization proposed by T.Beth and introduces a method for indirect trust calculation. To deliver recommendations in a secure and trustworthy manner rating certificates are proposed. Access rights of a peer are defined by trust in terms of restrictions and by contribution in terms of additional permissions.

The trust model proposed in [83] permits a mobile user to choose a service provider relying on the comparison of the trust value attributed to each known provider. The trust value is computed based on the satisfaction level, which is set by the user at the end of the session in a service provider's network. Interaction history is stored and computations are held at the Trusted Central Authority. In such a way, all users participate in the creation of reputation of each service provider. The need to communicate with the central trusted authority in order to obtain the trust value for a candidate network may affect user mobility.

A trust model for distributed environments, introduced in [84], uses direct trust, recommendation trust and role-based rules for making decisions about the trustworthiness of a peer. The main assumptions made in this model are that agents share their experience and agents are fair. These assumptions limit the possible implementation of the model to environments with established direct trust relationships.

The proposed models make use of complex analytical functions (such as beta function, exponential and logarithm functions) and the resulting trust value greatly depends on the parameters of these non-linear models. There is no direct and clear correspondence between values of model parameters and restrictions related to the peer's access policies.

III.6.2 Memory models

The basis of trust calculations is either a recommendation from the trusted third party or history-based personal observation of user behaviour. The appropriate model for retaining this history should be designed taking into consideration both the limitation set by the memory size dedicated to store history-related data and the timing factor. The timing factor is very important because more recent events must have more influence on the decision about the trustworthiness of the user, however the information about past behaviour, should also be taken into consideration.

Generally the history is represented by a sequence of single events, but in several works [99, 129,130,] history is represented by a sequence of finite sets of events.

The memory model based on a sliding window is used in [129, 130]. Since the log file is kept for a certain period of time the trust evaluation is based on a temporal factor. This sliding window implements a concept of consequent time units. One time unit includes a number of interactions (successful and unsuccessful). Once the time unit passes, the window slides from left to right, and the previous time unit is eliminated from the memory. Hence, the old experience history is not involved in trust computation.

This memory model does not keep track of the user's past behaviour and in case of dealing with a strategic attacker such a user will be considered as a trusted one following a time interval greater than the sliding window. Another problem is related to the choice of the window size. A small window size will lead to the fast "forgiving" of malicious users, while a large window size increases the amount of data to be processed and slows down the decision making.

The improvement of the memory model is proposed in [75, 130] by introducing a forgetting factor. A feedback or observation value from each party is calculated as a sum of the current feedback and the product of the forgetting factor and the previous feedback. In such a way an actual feedback value contains information about previous interactions. A disadvantage of this model is that the final feedback value does not depend on the time factor and recent information has the same value as old information.

Another approach to the user's behaviour history recording, based on fading memories, is proposed in the TrustGuard system [76] and used in [103]. It is assumed that the system stores a limited number of reputation-based trust values instead of results of interactions. Resent past observations are aggregated over time intervals. Using this model, the system needs to recalculate the recorded history after each interaction. Such a model prevents a malicious node from regaining trust quickly by keeping track of the previous bad behaviour but by the same reason the trust value for a malicious node decreases slowly if this node has shown good behaviour in the past.

K. Krukow, M. Nielsen, V. Sassone [85] propose an alternative way of recording behavioural information. Authors consider that information is lost when the information is recorded in an abstract manner [75,76,130]. A proposed policy-declarative language allows checking precise properties of trust behaviour. The general form of the proposed history recording includes indication of the recourse accessed, the time of access, time of validity and policy requirement. This model does not address the dependency between subsequent observation and the influence of very old observation.

III.7 CHAPTER SUMMARY

Security goals and objectives are the same for wired and wireless communications, but when applied to wireless communications more attention should be paid to confidentiality and integrity assurance due to the open nature of the air interface.

User's mobility nowadays is not limited to a home administrative domain, and therefore new authentication technologies are being developed. They take into account the more important handover characteristics: latency and security level keeping. The secure mobility management problem can be broken down into the following tasks:

- ✓ Visited network identity verification by the user;
- ✓ Dynamic trust establishment between administrative domains in a user-transparent manner;
- ✓ Secure communication between authentication servers;
- ✓ Fast user authentication in a visited domain;
- ✓ Fast (and ideally seamless) handover execution and
- ✓ Fast and secure redirection of the current session held with a corresponding node.

The security design becomes more challenging in the presence of mobility and flexibility becomes an important requirement along with CIA (confidentiality, integrity and availability) requirements. Even if very robust mechanisms are implemented for data confidentiality and integrity protection, these mechanisms are based on keys typically derived from authentication result. Thus, key material disclosure by a third party during a weak authentication may compromise overall communications held between authenticated entities.

Identity management system is based on one of the following trust models. Pairwise trust (direct agreements), brokered trust (agreements with a common intermediate entity) and community trust (agreements with a community or a federation). When responsibility is delegated to a service provider by an identity provider, the element necessary to add to the policy enforcement procedure is the history of previous interaction.

Traditional access control mechanisms are not suitable to the ubiquitous environment where all interacting entities are potentially unknown and therefore untrusted among each other. The number of users is extremely large and their behaviour is difficult to predict, so the risks for service providers change dramatically in such circumstances. It becomes impossible for an administrator to analyze system logs and adapt security policies to the actual situation. Thus a mechanism should be developed to provide access control to resources and to automated management of access policies. The concept of trust represents a promising basis for such a mechanism.

Chapter IV Towards secure ubiquitous networking

The today's digital universe is heterogeneous in various meanings of the word. Multiple IP-based services are offered for users that are subscribed to multiple identity- and service providers and have multiple roles. These users are equipped with multi-interface handheld devices with different capabilities and thus are able to access wide range of services over multiple access networks managed by multiple authorities. The limited scope of each access technology forces a user to gain connectivity through a variety of network technologies. For the same reasons different technologies coexist in the same geographical areas.

There is a strong need for new paradigms and approaches to manage this heterogeneous universe and to deliver to a user services adapted to his current terminal and access mode. Figure IV.1 represents the layered view of the communications architecture, and the management problems and solutions related to each layer.

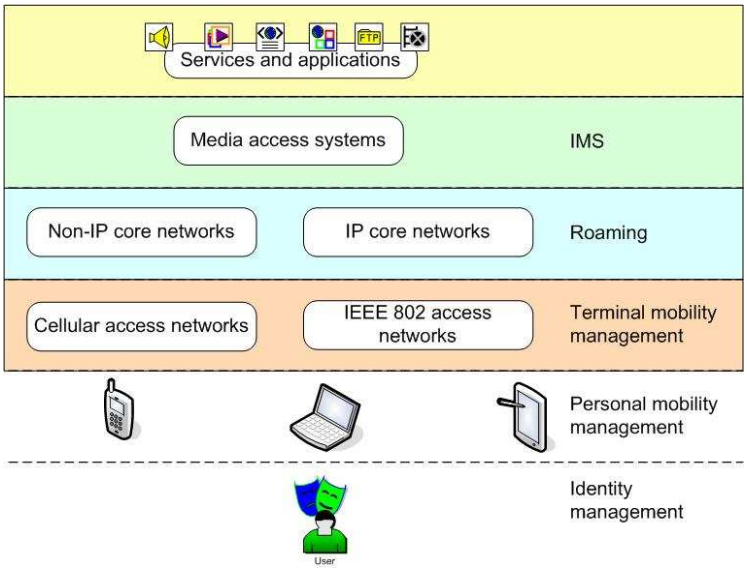


Figure IV.1: Multimedia services access

A wide variety of services are developed over the Internet; they range from non-real-time classic services such as web browsing, e-mailing and file downloading to new interactive services such as video transmission. Services traditionally proposed by telecommunication operators such as voce communication and messaging are migrating to IP networks. The global IP architecture that covers cellular and IEEE 802 operators will arrive soon. Ubiquitous mobility realization brings a lot of problems related to trust establishment, authentication and billing. These processes must be fast and reliable. As was discussed in Section III.4.1 , security threats are related to each handover step.

Numerous industrial, academic and standards institutions are involved in the work to overcome heterogeneity and to provide global, secure mobility management. This

chapter provides an analysis of key problems, existing solutions that address these problems, and issues associated with them.

IV.1 EXPECTATIONS AND EVERYDAY USE-CASES

A mobile user expects to move across networks seamlessly without knowledge about their structure, to carry out authentication only at the beginning of a session, not to have to care about protocols and services configurations (they should be self-configurable), to be assured of data protection and of the correct charge for services consumed.

At any time a user is able to choose between various access networks that coexist in the same geographical location. Networks have different coverage areas ranging from few square meters to hundred of kilometers, and they implement a wide variety of transport, routing and mobility management protocols. Services and signal strength provided vary in each network. Thus it is desirable to have a mechanism that realizes the "ABC" (Always Best Connected) concept allowing a user to choose the connectivity according to the current preferences and network applications running at the mobile terminal.

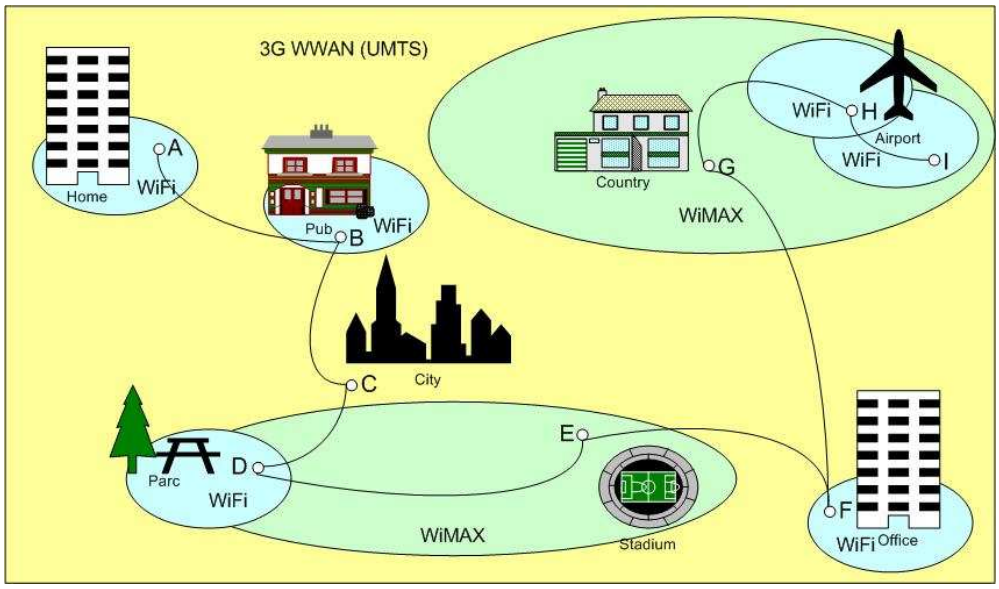


Figure IV.2: Everyday usage scenario

Figure IV.2 depicts an environment including urban and rural areas that a user visits daily or regularly. When being at point A, the user may choose between his home WiFi network and the network of his cellular operator. To start a voice call, the user selects a WiFi access due to the lower service cost. Continuing the call, the user leaves the home network and his terminal must automatically recognize signal strength degradation and reconnect to the same service via the cellular network, all transparently for a user. If the wireless network accessible at point B offers the required services, the user may choose to join it; otherwise he stays connected to the cellular network. Visiting the park (point D), the user starts a video-on-demand

session in a WiFi network and his terminal discovers the presence of WiMax network coverage providing better QoS and reconnects to it. The same session should be held when the user is moving from point D to G through points E and F. Being in the airport, the user starts a voice call in the WiFi network H and, changing an airport terminal (point I), seamlessly handovers to another WiFi network and finishes the call in it.

At each moment the mobile terminal should be able to choose the better access network and to decide to handover to it keeping the communication session active and not allowing the security level degradation. Network selection and handover initiation problems are addressed by the IEEE 802.21 working group [30]. IP-layer mobility is managed by Mobile IP means and session redirection to each new location is performed using SIP-based solutions. However most mobility management protocols were designed without inner security support and additional solutions need to be introduced in order to address threats presented in an open and mobile environment. Inasmuch as security signalling accounts for the overall handover latency, it has to be designed in the context of ubiquitous mobility management.

It should be noted that handover from an unlicensed network to a 3G network and vice versa are not symmetrical processes. Security specifications for IEEE 802.11 and 802.16 technologies allow the use of method-specific credentials and an extendable authentication protocol while 3G network authentication is typically operator-specific. If roaming agreements are established between a 3G operator and for example an authority managing a 802.11 access network, subscribers of the former can be authenticated in the partner's network using credentials provided by the home network. When roaming to a 3G operator's network a user must have appropriate credentials and his terminal must support SIM-based authentication.

IV.2 KEY CHALLENGES TO SECURE UBIQUITOUS MOBILITY IN A HETEROGENEOUS ENVIRONMENT

Analysis of security requirements for the mobility realization is a topic of work for many standard, academic and industrial institutions [86, 87, 88, 89].

This section provides a detailed analysis of security issues related to mobility support starting from candidate network selection and finishing with active session redirection to a new user's point of attachment.

IV.2.1 Network selection problem

Traditionally a network entity providing services advertises its capabilities as broadcast messages at the communication level corresponding to the service provided. An IEEE 802.11 access point advertises its name (ESSID) along with its security capabilities and provided bandwidth. If a user is subscribed to a cellular telecommunication operator, he knows from the contract what services are provided by his home network and by roaming partners of his home network. He has no possibility of finding such information in a dynamic manner.

The services provided by access networks vary from the simple link-layer connectivity to Internet connectivity and 3G operator-like services such as VoIP or

video-on-demand. The wide variety of services introduces a problem of "the best" access network selection by a mobile users. Moreover, a user should be able to find services as well as QoS, mobility and security support provided before being attached and authenticated to a network. Wireless access networks in the same geographical location may compete. For example, a WiFi network may advertise QoS support without really providing it. A mobile node choosing between a cellular and this WiFi networks may choose the latter due to the lower service cost for appropriate capabilities. After having connected to the chosen network the mobile user will encounter significant session QoS degradation. Another example is related to the use of an information service defined by IEEE 802.21. Such an information service provides mobile users with information about neighbouring networks capabilities. To believe in the correctness of information received a user must trust the source of information.

Another issue is related to the security of the network selection problem. Security must be provided for both of the access network and the mobile user. For a mobile user the candidate network discovery and selection problem has the following requirements:

1. Prevention of doubtful information, leading to a false handover or to a handover to a malicious network;
2. A mobile node should be able to verify integrity of the received information;
3. It is desirable to provide confidentiality of user traffic.

If a user obtains information from a non-trusted source this information cannot be considered as reliable.

Access network, it should be protected from DoS attacks and unauthorized access to internal entities.

IV.2.2 Security level maintenance and security matching

Security level maintenance is an important challenge for mobility. Access technology specifications and different access providers employing the same access technology may implement different security solutions for authentication, authorization, access control, key management, confidentiality and integrity protection. Table IV.1 illustrates possible security suits implemented in WWANs, WMAN sans WLANs. At the moment, no common metrics for security measurement and comparison exist.

Heterogeneity of credentials complicates the problem. When connected via a cellular network, a user is identified and authenticated by IMSI/TIMSI. In 802.11 or 802.16 networks user credentials include login/password pairs, digital certificates, tokens or public/private keys. In EAP authentication user identity is presented in the form of the Network Access Identifier (NAI) [90] that not only identifies a user to a network but also assists in routing the authentication request to his home network in case of authentication in a visited domain. Authentication methods and type of credentials should either be negotiated between the mobile node and the authenticator or they should be unified. The second approach does not seem to be realistic.

Two domains, establishing roaming agreements, should negotiate, adapt and translate their policies. Each user, arriving at a network, can have its own policies that also should be translated in a network format. A network needs to have the means for

dynamic policy adaptation to a user's needs. Most of existing policy management solutions is either platform or service specific.

Table IV.1: Cryptographic suits mapping

Wireless network / Characteristics	WLAN (802.11 example)	WMAN (802.16 example)	WWAN (UMTS example)
Authentication	802.1X/EAP	EAP/ X.509+RSA+SHA-1	UMTS AKA/GSM AKA
Key Management	802.11i [14]	PKM	UMTS AKA [55]
Keys	PTK/GTK	KEK/HMAC Uplink Key/ HMAC Downlink Key	Cipher key CK 128 bit, Integrity key IK 128 bit
Keys derived by	AS/AP - STA	BS	VLR/SGSN - ME
Integrity protection	TKIP MIC, CCMP	CMAC	CBC-MAC with KASUMI [91]
Data encryption	WEP/TKIP/CCMP	DES in CBC mode, AES in CCM/CTR mode	CFB/OFB with KASUMI Block cipher [91]

Security metrics should allow easy implementation in modern measurement systems. At least, if the user has started a session with confidentiality and integrity provided, the same security mechanisms should be provided in the new user's network of attachment.

IV.2.3 Dynamic trust establishment

In the mobile scenario there is no trust relationship established between the user and the visited network, thus they have to establish trust relationships dynamically.

Two networks may belong to the same service provider, be members of a permanent or spontaneous roaming coalition or may not have any relations at all. A user's authentication for a network is possible if he has either direct or indirect trust relationships with the authority managing this network. If a user is associated with a network, there is trust relationships established between them. When the user chooses a candidate network, the following types of roaming agreements are possible:

1. There are roaming agreements between the candidate and the mobile node's home network, but they are not defined for the candidate and the current network.
2. There are roaming agreements between the previous and the candidate network. The new network may or may not have trust relationships with the mobile node's home domain.
3. There is trust between neither the current and the candidate nor between the candidate and the home networks, however the mobile node has relations with the candidate network. These relationships may be based for example, on an independent subscription or pre-paid card.
4. The mobile user trusts a security broker that has trust relationships established with the candidate network.
5. No trust relationship is established.

The second and the third cases introduce a problem of indirect dynamic trust establishment. It is not guaranteed that trust establishment procedure will be successfully completed and that it will not be too long to interrupt a user's session. In the latter case handover to the candidate network is not possible.

A Public Key Infrastructure (PKI) was developed to provide the means for capabilities delegation. When dealing with scenarios where multiple administrative domains are involved, each administrative domain is served by a separate certification authority (CA) and surrounding PKI elements as it is shown in Figure IV.3. To accommodate roaming, these separate certification authorities must either conform to a hierarchy and a root certification authority or must provide cross-certifications. To avoid the problem of communication between different certification authorities and to make roaming agreement building less expensive, service providers may issue public/private keys themselves.

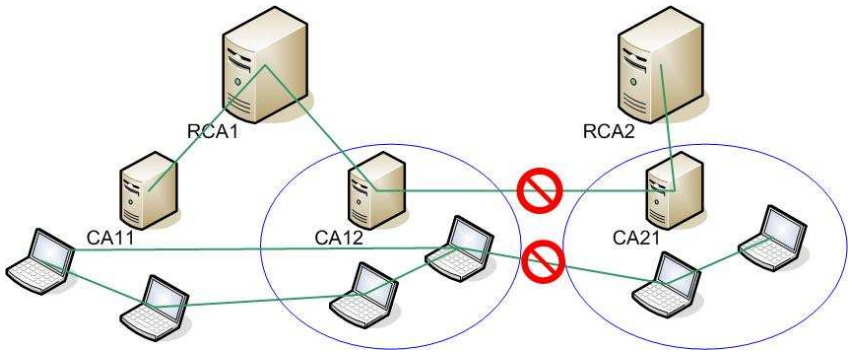


Figure IV.3: Interaction between Certification Authorities

Roaming agreements between administrative domains may change, security associations need to be renewed frequently and user lists are not static (users can change subscriber categories, require new services or change a payment mode). One organization may have hotspots in different geographical locations but with the same policies and subscribers. All these infrastructures need to be managed in a cost-efficient, secure and fast manner. Changes must be made in real-time mode.

Authentication represents a good mechanism for trust establishment. As was shown in Chapter III, strong mutual authentication has a significant impact on the overall handover latency, leading to QoS degradation and even to session interruption. Hence, authentication latency in an inter-domain and inter-technology scenario must be minimized.

Not only user authentication but also network authentication to a user becomes very important. Authentication with each access point is required, which affects handover performance.

IV.2.4 Fast mutual authentication

Access to different services is given to the user after he has been successfully authenticated in the access network, typically, without additional (separate) authentication for each service. That is why the strong authentication is required in this situation. Authentication must be also mutual to protect the user from a rogue point of attachment.

When a mobile node needs to re-affirm access to an authenticator or moves from one authenticator to another authenticator, the current authentication methods require a mobile node to execute a full authentication exchange with the authentication server in its home domain. Authentication time includes transmission delay between candidate and home networks and home authentication server processing delay. This conversation takes several round trips and significant time to complete causing delays in re-authentication and handover times.

It is undesirable to have to run a full authentication method each time a mobile node associates with a new authenticator or extends its current association with the same authenticator.

There are two fundamental approaches for authentication between the user terminal and the visited network. The first one requires communication between the visited and the home networks during the authentication process to verify user's credentials. These communications cause delays that are difficult to predict and to shorten. The second group of approaches is based on public key cryptography. Using X.509 certificates [92] may eliminate the necessity for inter-domain communication during the authentication process, but in this case authentication is only possible if both the mobile node and the visited network recognize each other's certification authorities. Disadvantages of the method are the heavy computational cost of asymmetric cryptography operations and the need for a certificate revocation mechanism.

In the mobile environment each security domain implements its own authentication mechanisms. Security solutions differ from one technology to another. Fast inter domain authentication must be independent of the technology of the serving and the current networks of attachment as well as of the method used for the previous authentication.

Security signalling during handover involves network access authentication and subsequent key negotiation in order to establish link-layer ciphering. Two communicating entities may trust ciphering and integrity protection keys only if they have established trust relationships. The strong confidentiality protection algorithm may be compromised if the key negotiation algorithm is weak. The security association between the user and the visited network must result from a strong mutual authentication process.

IV.2.5 User authorizations and access control in visited networks

Authorization-related issues are closely related to mobility. A visited domain should determine what a guest user is authorized to do in its network. Traditional access control is suitable for an environment, in which users that can associate with the network are known in advance and their rights are predefined. Home, enterprise and university networks represent this type of environment. In an ubiquitous mobility environment there is a priori a lack of trust among participants. This scenario assumes a number of participants, not known in advance, having (or even not having) different trust relations with each other. Operating in such an environment represents security risks both for the service providers' networks and for mobile users. Mutual authentication between the mobile user's device and the serving network ensures a certain security level that avoids communication with potentially dangerous partners. After verification of the network identity the mobile node can explicitly trust entities belonging to this network. On the other hand the fact that the network is able to

verify a user's identity does not mean that the network can trust this user. The number of users is much greater than the number of service providers. Users may be subscribed to different authorities and the presence of certain roaming agreements between these authorities allows users to be served by a non-home service provider. When making a decision about the trustworthiness of the client, the access network possesses only an acknowledgement of the fact that this client is subscribed to its roaming partner, obtained as the result of authentication. This acknowledgement serves as a recommendation for the user.

A user may demonstrate different behaviour in different networks. A trust model is needed to provide granulated access control to sensitive resources and to manage network security in a more effective manner.

IV.2.6 Secure redirection of a session with a corresponding node

After successful authentication in a new network a mobile node must inform its corresponding node or a remote service provider about his network address changing. This procedure presents a risk of DoS attacks, message forging, and false session redirection and so on. Thus this process must be secured. Location update and session redirection can be managed by two protocols: Mobile IP and SIP.

IV.3 EXISTING SOLUTIONS AND ASSOCIATED ISSUES

IV.3.1 Heterogeneous network IDs and user IDs management

If a network collaborates with or belongs to an operator, it may provide this information in a capability advertisement, as was discussed in the previous section. One operator may manage several access networks based on different access technologies. The user should be able to recognize the owner of an access network in order to provide correct credentials for authentication and to verify the network's identity.

As was mentioned above, a user who is subscribed to multiple service providers may have multiple identities. After a while, it becomes difficult for a user to manage and to remember all these identities and corresponding credentials. A number of efforts have been made to decrease the number of identities owned by a user.

If inter-provider federation exists, the Single Sign-On (SSO) concept allows a user to access different domains while being subscribed to only one of them. Open ID [93] proposes a way to use a single digital identity over the Internet. Liberty Alliance [94] and Open ID allow connection to different sites with the same identity. Shibboleth [95] is a project that proposes an architecture for federated identity-based authentication and authorization. A federated identity allows a user from one security domain to be served by another security domain that belongs to the same federation. Identity providers create user identities and service providers use them to provide service access. To enable user authentication and authorization by a service provider, the latter must belong to the same federation as that the user's Identity provider and they must have service level agreements and security association established. Authentication between a service provider and a user is either performed by authentication request redirection to the user's identity provider or is based on

cryptographic cookies valid not only at the issuer server but also at its roaming partners.

These approaches save users from remembering multiple identities and passwords, but they do not provide physical mobility, only "virtual" mobility, which means that a user can visit web sites of different service providers without the need to be authenticated while connecting to a new site.

IV.3.2 Security mapping

L. Reznik has provided an analysis of existing methodologies of computer system security and information assurance measurement [96] such as "Orange Book". He concludes that these systems are too complex to be implemented for dynamic security evaluation for modern computer systems. Methods for security evaluation were defined for program codes and static systems but we cannot say unambiguously that one authentication protocol is more secure than another.

At the mobile user's side the possible solution to match security provided by the serving (home) and candidate networks is to explicitly define security preferences. These security preferences may include a list of supported authentication protocols, keys lengths and encryption and integrity protection algorithms. Such security profiles may be defined for each type of service that a user may require.

When a mobile user is served by a non-home network at least three groups of security policies should correspond: the user's policies, his home network's policies and the visiting network's policies. IETF RFC 4017 [97] defines requirements for EAP methods used in wireless LAN authentication. Mandatory requirements listed in this document are also addressed by WiMax standard [3].

The use of the trust concept to establish trust relationships in federations has been proposed in academic work [98, 99, 100, 101]. The main purpose of such models is to negotiate access policies applied to subscribers of one domain served by another administrative domain. The possibility of indirect trust relationships among previously unknown entities is studied in [102]. A system for trust management between different peer-to-peer communities that involves only delegated peers (leaders, the most trusted entities in the group) in this process in order to decrease network traffic is proposed in [103]. In other models a peer computes the trust value from feedbacks received from all peers in the network.

IV.3.3 Secure network selection and handover decision

To obtain information advertised by an access router, a mobile node must associate with the network, to which this access router belongs and obtain an IP address. In such a way a user is not aware of services provided by the candidate network before associating with it.

Pre-authentication signalling proposed in [104] allows a mobile node to learn WLAN service capabilities and costs before association. The use of the IEEE 802.1X framework is proposed to provide capability information to clients before authentication. The possibility of secure data exchange is based on the use of the security association between the candidate and the user's home service providers. The candidate network discovery and selection problem in heterogeneous environments is addressed by the IEEE 802.21 working group, which introduces information services providing information about neighbouring networks of various access media.

However, the security aspect was not addressed from the very beginning of the concept design. The 802.21 Security Task group addresses the security of Media Independent Handover services. The approach includes authentication of the information server to the mobile node before information exchange and the mechanisms for secure transport of information definition. The ongoing work on the MIH Services security design, in which we are taking part, is presented in Annex C. The *Ambient Networks* European project [105] aims to develop a user-centric framework providing service delivery independently of the access technology used. The framework is intended to operate over a variety of infrastructures. The project works on mobility management, inter-working, multi-access and service aware transport issues. Along with that Ambient Networks project addresses the problem of how to prove correctness of information advertised by a network before association. The proposed solution is based on capability advertisements signed by a trusted third party. A well-known and widely trusted authority signs public keys of access providers. A network access provider signs advertised network capabilities with its private key and includes in each advertisement message its public key signed by the trusted third party.

IV.3.4 Dynamic trust establishment and trust delegation

A mobile user is able to associate with an access network managed either by an authority to which a user is subscribed or by another authority that has established corresponding agreements with the user's home operator. A user and a visited network establish indirect trust relationships via the common trusted third party.

Bilateral trust relationships establishment is not a scalable solution for the ubiquitous mobility scenario because with an increasing number of collaborating service providers the number of relationships and corresponding security associations to manage increases dramatically. N service providers must establish $\frac{N(N-1)}{2}$ associations. *Brokered relationship* helps to decrease the number of associations to manage. In this case users trust a security broker also trusted by various service providers. For N service providers only $N-1$ associations should be established. The presence of a single point of failure is the main weakness of this approach. Compromising the security broker leads to violation of security of communication of all its partners.

Federations of service and identity providers represent a case of multilateral trust relationships. Each operator has multi-way relationship with several members within an alliance or a federation. A new member may build trust relationship with a few members initially or on an as-needed basis (based on user service requests). Trust relationship building in real-time introduces a performance issue. Usually, a user subscribed to one federation is not able to seamlessly handover to a network belonging to another federation.

Trust establishment between the mobile node and the candidate network requires information exchange between the candidate network and the authority trusted by a user. As this exchange causes significant latencies it is desirable to enable the trusted third party to delegate trust for its users to its partners. The use of cryptographic cookies understood not only by its issuer but also by all members of a federation, and digital certificates are examples of trust delegation.

The use of self-assigned public-key based identities has been proposed in the *Ambient Networks project* [105] in order to eliminate the need for security infrastructure to bind identities to public keys.

Authors of [62] define trust negotiation as an emerging access control approach that aims to establish trust between negotiating parties online through bilateral credentials disclosure. This work addresses the challenge of distributing the functionality between identity and service providers and discusses how to integrate identity management with trust-negotiation techniques. They propose an integrated identity management and trust negotiation.

IV.3.5 Fast authentication and handover performance

IV.3.5.1 Intra-domain handover

When a user roams within one administrative domain, only the first authentication between him and a network is full. All following authentications may use fast methods. Numerous approaches for fast re-authentication were proposed in unlicensed networks. Some of them are technology specific while others are designed to be technology-independent.

Handover support in cellular networks

Cellular networks support user terminal handover between base stations. With the coming of 3G systems one operator can manage both 3G and 2G access networks. That is why interoperability between them should be provided. As 2G entities are not capable to proceed within the 3G security context, converting functions are performed by 3G entities. The interoperability between standards is explicitly defined in UMTS Security Architecture Technical standard [55].

The solution for the inter-technology case of the intra-domain roaming is already proposed and implemented by telecom operators – *Unlicensed Mobile Access* (UMA) [106]. The subscriber of a telecom operator is able to access GSM and GPRS mobile services being authenticated in an unlicensed network such as IEEE 802.11 or 802.15 managed either by this operator or by its partner. Usually the implemented authentication method is EAP-AKA [107] or EAP-SIM [108]. UMA architecture consists of an access point and one or more UMA Network Controllers (UNC) interconnected through a broadband IP network. To access the operator's services the mobile node communicates with UNC. Security mechanisms implemented in UMA include security of an unlicensed network, security mechanisms protecting signalling, voice and data traffic over the interface between the mobile node and UNC and including both authentication and encryption. The mobile node needs to perform authentication with the HLR in the operator's network. The handover between access networks is announced to be transparent for a user. Thus to access services a subscriber must perform three subsequent authentications. Simultaneous use of two network interfaces makes possible soft handover execution and that is why the handover process is transparent to a user. Location services allow session redirection to a current user point of attachment while he is performing authentication process.

Fast authentication support in IEEE 802.16

The IEEE 802.16e standard [3] specifies that a mobile node may perform handover from one base station to another. A handover management scheme may be localized at base stations and it may involve Authentication and Service Authorization servers.

In anticipation of a handover, a mobile station may seek to use pre-authentication to facilitate an accelerated re-entry at a particular target base station. Pre-authentication results in establishing of an Authorization key on the mobile station and on the target base station. The specific mechanism for pre-authentication is not yet defined.

Fast authentication in IEEE 802.11 networks

802.1X authentication provides a high level of security but accounts for the majority of overall handover latency. Some methods to reduce layer 2 authentication delays are proposed. Fast authentication methods allow the mobile node to establish a trust relationship with a new access point via a relationship with an old access point. Instead of the keys used by an old access point for encryption a result of some function of these keys is distributed. The proof of the client's identity is encapsulated in the EAP Identity Response.

Pre-authentication defined in the IEEE 802.11i amendment [14] supports one-to-many message exchange. Mobile node associates and authenticates with an access point. When being associated the mobile node executes active or passive scanning and selects one or more candidate access points. Then the mobile node authenticates to the candidate access points before connectivity is lost with the serving access point. To perform pre-authentication, unicast data frames are sent to the candidate access point and the serving access point forwards them. Finally, the mobile node handovers and associates to the one of the candidate access points. This approach facilitates a considerable decrease in authentication latency but it also opens up new possibilities for DoS attacks and causes significant traffic overhead. A mobile node can be pre-authenticated only in the same LAN within the same access router.

Predictive authentication proposed by S. Pack and Y. Choi uses a modified 802.1X key distribution protocol [109]. One-to–many authentication is performed instead of one-to-one authentication. The mobile node sends an authentication request to an authentication server via one access point and after successful accomplishment the authentication server sends key material to all access points in a region. The significant network load can be reduced by choosing a *Fast Handover Region (FHR),* the set of access points that are most frequently visited by the mobile node. A FHR is commonly described with Neighbour Graph (NG) [110], which is used to determine the candidate set of access points with which the mobile node could potentially associate. This dynamic data structure can be maintained in a distributed manner among access points or in a centralized manner at an authentication server. The implementation of this scheme requires construction of mobility pattern for each user and that is why it cannot be implemented in open public access networks.

Proactive Key Distribution (PKD) proposed by A. Mishra in [111] enables a reduction in handover latency and the use of a mobiles station's computational power by pre-distributing key material (PMK) ahead of the mobile node. Enhancements to the Proactive Key Distribution were introduced by M. Kassab, A. Belghith, J.-M. Bonnin and S. Sassi in [112]. They have reduced the duration authentication exchange between a station and a target network by providing the station and neighbour access

points with PTK via two methods: IAPP caching and anticipated 4-way handshake. The proactive method requires changes at the client, access points and an authentication. Instead of a four-way handshake a two-way handshake is used after key distribution. The disadvantages of the method include the requirement for new RADIUS messages and network load increasing. The handover latency depends on many parameters such as signalling overhead, the number of hops between an access point and an Authentication Server and the number of access points supporting pre-authentication. To evaluate the influence of these factors on the handover performance, an analytical study has been carried out in [113].

Reactive method for key distribution [111], unlike Proactive key distribution, does not require changes at an access point. When an access point sends an EAP Identity Request to a client, it responds with a PMK identifier, the access point asks an authentication server for a corresponding key (which results in two messages), it sends an EAP Success message and a four-way handshake is performed. The old PMK must immediately be deleted and a new one generated.

Another method to reduce authentication latency proposes the use of IEEE 802.11f [27] protocol for *secure context transfer* [114]. 802.11f (Inter Access Point Protocol - IAPP) is designed to exchange context between a current access point and a new one during the handover process but it provides a standard set of messages, which can contain security-related information. *Direct L2 context transfer* for inter-ESS handover requires a roaming server, roaming agreements and NAT traversal mechanism. This solution can be implemented only if the access point has a public IP address. *Encapsulation of L2 context in L3 context* presents an integration of Context Transfer Protocol (CTP) [115] by the Seamoby working group, and IAPP.

N.Aboudagga, M.Eltoweissy and J.-J.Quisquater developed an approach for fast authentication [116] for roaming in the same domain and the same WLAN subnet. The ticketing scheme is proposed to reduce authentication latency and dependability on the server by pre-distribution the keys to the access points in the mobility pattern of the mobile node. The tickets are distributed between access points using IAPP. The second proposed scheme introduces authentication tokens. The mobile node uses a token to authenticate with a candidate access point without interaction with the server.

The above-listed methods focus on the horizontal handover within the same administrative domain. They allow maintaining a high security level but they require lengthy observation, logging and analyzing of a mobile node's behaviour. They also cause network load increases. These factors restrict the possibility of implementation to office/enterprise networks, where there is a constant set of users with stable mobility patterns. Networks, open to public access, cannot grant the required amount of computational resources and traffic to visitors.

IV.3.5.2 Technology-independent fast authentication methods

The *IETF Hokey* working group [117] proposes two solutions for fast intra-domain authentication: *pre-authentication* and *hierarchy-based* authentication. Hokey implements a generic mechanism to re-use derived EAP keying material for handover. The EAP hierarchy defines two keys that are derived at the top layer, the master session key (MSK) and the extended MSK (EMSK). As the MSK is used for session keys derivation, it is proposed to be used as a top of the handover key hierarchy.

Another goal of the work is to provide independence from the lower layers, in order to assure protocol operation in the heterogeneous environment. Key transport between different entities involved in the handover keying architecture may require additional transport protocols. In the case of pre-authentication two scenarios are possible: on-air authentication when the mobile node is able to authenticate with the target access point being attached to the serving access point. In this case the concurrent link-layer attachment must be supported. M. Nakhjiri has shown in [118] that the handover keying key hierarchy may be successfully implemented for secure handovers in WiMax.

Another scenario proposes inter-DS signalling, when the mobile node communicates with the target access point via the serving access point. Several problems remain: the mobile node must be able to discover the address of the candidate access point (either L3 or L2). Normally an access point has an internal non-routable IP address. IEEE 802.11f [27] (inter-access point interaction at L2) and IETF Seamoby [119] (inter-access router interaction at L3) propose remedies for this issue. An analysis of mobility optimizations for PANA such as pre-authentication and context transfer is provided in [120] and it was concluded that these solutions are not suitable for inter-domain handovers due to their high cost and the need of trust relationships and security association established between PANA Agents.

IV.3.5.3 Inter-domain fast authentication solutions

There are two possibilities for a network and a user to authenticate each other. The first one is based on certificates and the use of public key cryptography. In this case it is not necessary to communicate with a mobile node's authentication server to complete an authentication, if both the mobile node and a visited network can recognize each other's certification authority. Nevertheless, the home authentication server must be notified to start the user's accounting in the current network. The second authentication approach requires communication between foreign and home authentication servers to authenticate a user. User's identity should include a link on its home network. Network authentication by a user is an open issue in this case.

Ambient networks project [105] proposes two authentication mechanisms: the one introduced in Host Identity Protocol [39] and Diffie-Hellman key agreement based on self-signed public key certificates.

The IEEE 802.21 Security Task Group addresses problems of security signalling optimization during handover in inter-technology, intra- and inter-domain scenarios. EAP has been chosen as a candidate protocol for carrying authentication messages. For intra-domain handover, a key hierarchy-based solution proposed by HOKEY [117] has been adopted, while during inter-domain handover a mobile node should use a pre-authentication based solution. However, a lot of performance, scalability and security issues are related to this approach. Detailed analysis of these issues is provided in Annex C.

L2 and L3 based authentication

Several networks allow users to access the internal entities without the link-layer authentication. The problem of choosing the level on which the authentication is performed is specific to 802.11 networks since mandatory authentication with an access point is implied in technologies such as WiMax and UMTS.

IEEE 802.11i, Web authentication and PANA [121] are commonly used hotspot authentication approaches today. The first one requires authentication with an access point, others allow network access to any user, however traffic from unauthorized users is filtered by a gateway device.

Two network entities can play an authenticator role: an access point and an access router. 802.1X framework provides strong mutual authentication between the mobile node and the network authentication server via an access point. A lot of fast authentication approaches are designed to exchange user's identity information and key material between access points and authentication server in the same administrative domain. When there is a need to extend fast authentication methods for inter-domain operations, access points involving causes many difficulties. Access router may play a role of authenticator in PANA [122], while user pre-authentication is achieved by means of context transfer protocol proposed by Seamoby group [119]. Table IV.2 summarizes advantages and disadvantages of these two cases. Disadvantages are marked in grey.

Table IV.2: Access point and access router suitability for the role of authenticator

Access Point	Access Router
AP is the first network entity which a MN communicates with;	To communicate with an AR, a MN should have a priori knowledge about it;
AP is responsible for encryption key negotiation with a MN;	AR usually acts in link-layer independent manner, thus it does not provide low-layer keys established;
AP usually has an internal non-routable IP address, thus all messages will pass through the AR;	AR always has a global and an internal IP addresses;
To provide pre-authentication between different networks via APs, each network must present to another a detailed description of its topology;	A network always presents information about its ARs;
Multiple APs must be configured, this procedure is excess;	A single AR configuration is required per each subnet;

Authentication with an access point protects all internal entities from unauthorized use while authentication with an access router opens up a possibility for different types of attack at internal network nodes: on access points, DHCP server etc. Application-layer authentication represents the same risks.

A compound layer-2 and Web authentication scheme is proposed in [131] to ensure cryptographically protected access in public wireless LANs. According to this scheme, the user first establishes an L2 session key by using 802.1X guest authentication. After that he embeds an L2 session key digest in the web authentication. Guest access to the network may cause a security problem: an unauthenticated user can monitor a wireless channel, acquire an IP address and perform DoS attacks against network entities and authenticated mobile nodes. In addition, time taken by Web authentication often does not permit a real-time application to continue running.

Inter-domain extensions of intra-domain fast authentication methods

Fast authentication methods designed for intra-domain handover have shown good results and there is an attractive possibility to use them for a case of inter-domain roaming. *Straightforward extension of IAPP* assumes that there are trust relationships

between old and new visited domains and between access points of two domains, and that a pre-existing IPSec or TLS tunnel exists. The mobile node's home authentication does not participate in authentication; a new domain should accept all users coming from the old domain, but it cannot verify a chain of trust if it has no roaming agreements with the mobile node's home domain.

Inter-domain proactive key distribution pro-actively transports PMKs between handover candidates in different networks. This approach involves the home authentication server, which is responsible for new PMK creation. The mobile node's home authentication server must have information about the topology of different domains to compute a set of possible APs close to a current location of the mobile node. Full responsibility is placed on the home authentication server. This fact can cause a significant authentication delay due to round-trip time.

Pre-authentication over multiply domains requires two access points that known and trust each other, a pre-established secure link between them. Every access point needs knowledge about nearest access points and the mobile node itself needs information about possible access points, with which it wishes associate. Authentication is governed only by roaming agreements.

To optimize security signalling for the inter-domain handover EAP pre-authentication has been proposed in [86]. Two modes of EAP pre-authentication are defined: direct pre-authentication and indirect pre-authentication. A study of this approach's applicability to inter-domain handover is provided in Annex C.

Ticket-based authentication methods

There is a strong trend to use mechanisms described by the Kerberos protocol [123] for inter-domain authentication because they eliminates need for communication between the local authentication server and a remote server to verify a user's identity and provide means for trust delegation. Another direction in security-related handover signalling that is represented in academic work aims to minimize communication between the mobile node and visited network with the mobile node's home network.

Z.Hong et al proposed a "Passport Based Fast Authentication" (PBA) [124]. Authentication servers of networks that have roaming agreements create a common shared key to authenticate a user in a visited network. The home network provides its users with a so-called passport that allows the visited network to authenticate a user without consulting the home network. The passport contains an identity of the mobile user and is signed with a common shared key. The passport revocation mechanism requires communication between the user's home network and all its trusted networks. Since all trusted networks share the same key, the user's passport can be signed by both the home and the visited authority. This point represents a management issue: the validity period of the passport signed by the visited network may exceed the term of the contract between the user and its home network. This enables authentication of the invalid user in other visited networks. The approach is suited to the federation of networks with multilateral roaming agreements but it is difficult to create a shared key in case of bilateral trust. If a new authority enters the federation, the shared key must be recalculated and all users' passports must be refreshed.

The work of H.Wang and A.R.Prasad on fast authentication for inter-domain roaming [125] introduces the concept that reduces the authentication latency due to a low round-trip time of communication between geographically neighbouring networks. The serving network generates a keying material based on the current security context and delivers them to both the mobile node and the target network. This information allows mutual authentication between the target network and the user. The method starts following a handover decision made by the mobile node, serving network or both of them. If the handover happens before the serving network makes a decision about it, the reactive scheme is implemented. In this case the network delivers keying materials to the user just after successful authentication, and the target network asks the serving network for the keying material after the user authentication request. The approach proposes a network-driven scenario for authentication, and its implementation is only possible if there are trust relationships and a protected channel between the serving and the target networks.

M.Long, Ch.-H. "John" Wu, J.D.Irwin proposed a Localized Authentication for Wireless LAN Inter-network Roaming [126], which enables the candidate network to avoid communication with the home networks to authenticate a user. The approach is based on the public key cryptography and adapts the SSL handshake protocol. Each network operator represents a certification authority and issues certificates for its subscribers. Each authority keeps its certificates signed by all roaming partners to facilitate network authentication by the visitor. The user has public keys of all partners of his home network. The visited network verifies the user's certificate signed by the roaming partner and encrypted with the symmetric session key, generated during authentication exchange. The approach requires heavy management of credentials, while public key cryptography operations cause high authentication latency.

S. G. Polito and H.Schulzrinne [127] introduce a consortium-based trust model among providers to allow shared authentication and authorization of their users. The service provider that has a contract with the user is responsible for paying other service providers for services consumed by a user. The model supposes one private certification authority per consortium. Each service provider issues authentication and authorization tokens for users and allows other service providers belonging to the same consortium to obtain and to verify these tokens. Tokens serve to authenticate users and to indicate their authorization profiles. The certification authority is responsible for providing each service provider with a certificate to let them be authenticated by users. The provision of the token is dynamic within the consortium and each service provider may ask for a token the service provider previously visited by the user instead of this home service provider. It allows reduction of the authentication delay. The token is also used for call authorization until its expiration.

Y. Ohba introduces a scheme for media-independent handover key management architecture based on the use of the Kerberos protocol [128]. The mobile node is able to obtain a master session key required for dynamic security association establishment with an authenticator and the authentication server without communicating with them before a handover. Modifications to Kerberos protocol concern relationships between the Key Distribution Center (KDC), the Authentication Server (AS) and the user. All entities may be managed by different authorities, but there must be trust established between the AS and KDC and the user and KDC. The mobile node first requests a ticket granting ticket from a Kerberos Distribution Center. When the

mobile node discovers one or more candidate authenticators it asks for tickets for this authenticator using the received TGT. After that the mobile node makes a handover for one of the chosen authenticators. If the mobile node handovers to another authenticator, it executes authentication in reactive mode. In this mode, the roles of the authenticator and the mobile node are inversed: the mobile node is a server and the authenticator is a client. It should be mentioned that there is a possibility of a DoS attack in a reactive mode of operation since the trigger message is not protected. An additional mechanism should be provided to mitigate DoS attacks.

IV.3.6 Access control in open environments

In a dynamic and heterogeneous wireless universe, access control becomes very problematic. To determine what actions a mobile user is authorized to perform in a non-home network the use of service level agreements with a user's home network and authorization or attribute certificates is proposed in [105, 127, 94,95]. Access control policies should be based not only on authorizations delegated by trusted parties, but they should also take history and context into consideration.

Approaches for trust-based access control (TBAC) [100, 129, 130] propose a two-step operation for granting access to a user. Firstly, the system computes trust for a particular client and after that an access role is associated with a user according to the trust value computed in the previous step. Trust calculation is based either on recommendations or on observations. When information collected on the particular client is not sufficient for decision-making, a recommendation from the trusted third party is used [129].

IV.4 CHAPTER SUMMARY

In this chapter we demonstrate our vision of the future wireless networks that support ubiquitous secure mobility. The main challenges introduced by ubiquitous mobility are: scalability, interoperability, QoS guarantees during handover and security.

Mobility management approaches proposed have demonstrated good performance but handover latency is still affected by security-related signalling. The simultaneous need for strong security providing and handover latency minimization introduces a complex research problem. The main issue is related to the authentication between the mobile user and the visited network because all confidentiality and integrity protection mechanisms to be implemented depend on the authentication result. Thus to protect the network from unauthorized access, user authentication should be performed as early as possible, preferable at the link-layer. Inter-technology handover presents the same challenges as inter-domain moving, but adds a requirement that all solutions be link-layer independent. Service discovery procedure must not come up against network security. Trust establishment between networks of different organization is an open issue.

There are many limitations for security solutions deployment in the mobile environment. First of all, the authentication and confidentiality establishment procedures must not significantly increase the handover latency. To achieve this, we

need to avoid inter-domain signalling during the handover execution by replacing them with the post-authentication signalling. All authentication solutions should not implement heavy computations due to the short battery life and limited time. The problems with carrying it out may be classified as technical and non-technical (business).

Numerous solutions have been proposed to optimize security signalling. Intra-domain fast authentication is a problem that is almost solved, while inter-domain authentication remains very challenging. Fast authentication methods that modify the standard have shown good results for intra-domain handovers and represent an attractive possibility to use them for inter-domain roaming. But such extensions of proposed approaches require the establishment of trust relationships between internal entities of different networks, such as access points or access routers. Using access routers as authenticators makes the authentication technology-independent, but opens access to the network at the link-layer for all potential clients.

Section IV.3.5.3 provides an overview of fast authentication approaches proposed in literature and their principle characteristics. Two groups of fast authentication approaches for inter-domain handover have being recently proposed and developed: pre-authentication and ticket-based authentication. Pre-authentication is inspired by the possibility of a mobile node to start authentication exchange prior to association defined in 802.11i and 802.16m standards. The need of trust relationships established and possibility of communications between the current and candidate authenticator limit their implementation to one administrative domain. Ticket-based schemes introduce more promising approach for the multi-provider and multi-federation environment.

Fast re-authentication approaches introduced by M. Long [126] and S. Polito [127] assume the presence of a certification authority that issues certificates for federation members and helps to manage user public keys. The format of certificates can differ and certificates issued by members of one federation are not understood by members of another federation. For successful operation of the protocol proposed by Z. Hong [124] all federation members must share a single key. If a new member enters the federation, all issued tokens must be re-calculated and re-distributed. H.Wang et al [125] do not make assumptions on the security associations but their proposal requires inter-domain communication during authentication and key material delivery to both the target network and the mobile node. The Kerberized handover Keying introduced by Y.Ohba [128] shows good performance in proactive operation mode. It is assumed that the client and the server have pre-established trust relationships with a Key Distribution Center.

In cited approaches assumptions about the nature of security associations between roaming partners are made. The scenario where security associations with different partners are based on different cryptography types is also possible. The user's home provider can have bilateral and multilateral roaming agreements with multiple service providers. For example, a service provider is a member of federation where all members share a single key; it belongs to another federation having its own certification authority that issues public key certificates for federation members. Bilateral security associations are based on either symmetric or asymmetric cryptography.

As each proposed fast authentication approach assumes a particular nature of roaming agreements and security associations established between the partners, each partner needs to implement multiple fast-re-authentication solutions in order to allow subscribers seamless roaming across all its partners. The client must adopt multiple fast re-authentication protocols and it have to manage a huge amount of information such as public keys of partners belonging to different federations or other associated credentials. This solution ensures the possibility of fast authentication across all roaming partners of the user's home provider, but it seems to be redundant, non-efficient and non-scalable.

To provide a mobile node with a possibility to perform fast authentication with all partners of its home network, there is a need for a fast authentication solution that is independent from the nature of security association between partners. The particular implementation of these associations must not have impact of the process of credentials issue for a client.

The SSO approaches described are not suitable for terminal mobility management for several reasons. First of all, the authentication signalling causes unacceptable latency for handover execution (up to 2 seconds according to measurements done by Y. Matsunaga [131]). Secondly, authentication is performed at the application layer, in other words, after the user terminal has obtained access to a network.

Table IV.3: Security requirements, existing solutions and associated issues

No	Requirement	Solution	Issues
1	Secure network selection	Ambient Networks IEEE 802.21	Unauthenticated information
3	Dynamic trust establishment between previously unknown parties	Mutual authentication methods	Authentication exchange causes unpredictable latencies
4	Overcome heterogeneity of credentials	Open issue	
5	Fast authentication in inter-domain handover	Pre-authentication, Ticket-based schemes	Authentication takes a long time; pre-authentication introduces significant scalability, interoperability and security issues. Inter-domain communication is required for authentication. Strong assumptions on the nature of security associations between roaming partners
6	Access control in open environments	RBAC TBAC	Difficult to adapt policies to continuously changing dynamic environment, unknown in advance users and unstable user access lists

With the approaches proposed the mobility of a user is limited by service providers' networks that have roaming agreements with his home service/identity provider. Many models have been proposed for trust based access control but they do not cover

all requirements introduced by the modern environment for mobile access to services. The previously proposed models do not set direct and clear dependency between the risk level in the managed environment, network access policies and the parameters of trust computation model.

Another issue related to fast authentication is the fact that in some architectures such as Ambient Networks [105] and IMS [9] authentication to the access network is separated from authentication to a service.

Table IV.3 summarizes open issues related to secure inter-domain and inter-technology handover. In this thesis we concentrate on the problem of fast mutual authentication and flexible access control in the inter-domain handover scenario.

In order to decrease the handover latency caused by security signalling the authentication process may be broken down into pre-authentication signalling and fast authentication during a handover process. The authentication method must be independent of transport, technology, the authentication method used previously and it must provide key material for future key generation. EAP is a good candidate to carry fast authentication messages because EAP is extensible, it is mode and media independent and it is used in the 802.11, 802.16 and 3GPP standards.

Traditional security mechanisms are based on static agreements while today's digital communications require more flexible and autonomous security solutions reflecting dynamically changing trust relationships among partners.

Chapter V Fast inter-domain authentication

This chapter describes our solutions that address reducing authentication latency during inter-domain handover and security-related signalling optimization.

We propose a compound method for user authentication in a public access wireless LAN when the latter requires separate authorization to access internal network services and the Internet. The approach we develop aims to minimize a risk of attacks at network nodes conducted by unauthenticated users provides key establishment and strong encryption between a mobile node and an access point and decreases overall handover latency. An authorized user is granted network and Internet access as a result of a single authentication process that combines 802.1X and PANA operations. This work is focused on reduction of user authentication time in a visited WLAN, protection of visited network's internal entities and negotiation of user's encryption key, all in a single process.

We then introduce the Fast re-Authentication Protocol (FAP) for inter-domain roaming, which aims to reduce authentication delay of a mobile user in a visited administrative domain. The approach eliminates the need for communication between the target and the user's home networks for credentials verification, and uses a short-living lightweight re-authentication ticket that does not require revocation mechanism. The proposed approach does not depend on the nature of roaming agreements between different networks. To minimize the number of authentication tickets sent to each subscriber, we propose a ticket distribution scheme.

V.1 COMPOUND USER AUTHENTICATION TO A WIRELESS LAN: THE FIRST STEP TO HANDOVER OPTIMIZATION

V.1.1 Purpose of the work

PANA is a protocol for a mobile node's authentication to a first access router. It serves to transport EAP packets over an IP network and does not depend on a link-layer carrier. Authentication with an access point protects all internal entities from unauthorized use while authentication with an access router opens up the possibility for different types of attack on internal network nodes: on access points, DHCP server etc. Web authentication presents the same risks.

A compound layer-2 and Web authentication scheme is proposed in [132] to ensure cryptographically protected access in public wireless LANs. According to this scheme, the user first establishes an L2 session key by using 802.1X guest authentication. After that he embeds an L2 session key digest in the web authentication. Guest access to the network may cause a security problem: an unauthenticated user can monitor a wireless channel, acquire an IP address and perform DoS attacks against network entities and authenticated mobile nodes. In

addition, time taken by Web authentication often does not permit a real-time application to continue running.

A network can propose different types of services. Some authenticated users need only to have Internet access to continue a session, others need to use internal network services. Such a scenario can require a separate user's authentication between a link-layer connectivity provider and an Internet service provider [133].

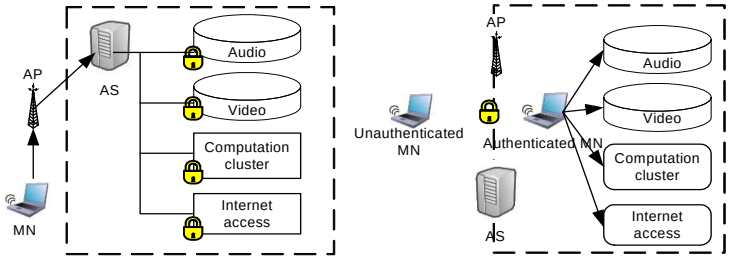


Figure V.1: Types of user access to services provided by a network

A mobile user arrives in a new network intending to continue a real-time session with a certain corresponding node. Figure V.1 depicts two types of network services access: network-managed and user-managed. Several scenarios for user access to services are possible:

1. The mobile node authenticates with the authentication server via an access point using 802.1X and after that it has access to all services on the network;
2. The mobile node authenticates with the authentication server via an access point and must be authenticated to get access to each service and
3. The mobile node does not authenticate (or performs guest authentication) with an access point and must be authenticated to a network access server (the case of PANA use).

The majority of mobile users need access to an access router to communicate with external networks. To prevent unauthorized network usage, the access router must know the user's identity. There are various ways to achieve this:

1. When the mobile node authenticates with an access point, the latter transmits its identity-related information to an access router;
2. The mobile node must execute authentication with an access router itself; and
3. Authentication with the access point and the access router is done at the same time.

For real-time applications the main requirement is that as little time as possible should be taken to change a point of attachment. So there is a need to combine authentication of an access point and authorization to a service (an Internet service provider) in a single process. Other services do not require transparency.

Our work focuses on the first full user authentication in a new administrative domain, but fast authentication methods for subsequent cell and subnet handovers; such as for example 802.11i predictive authentication [134, 112] and PANA mobility optimization [135] might be implemented.

V.1.2 Model and assumptions

Certain networks may offer access to a limited topology (link-layer connectivity and limited network layer access) for unauthenticated visitors, but any access beyond this topology requires authentication and authorization [133].

To communicate with a PANA Authentication Agent (PAA), a mobile node must have an IP address. The PANA draft [121] assumes that a user can have different addresses before and after his authentication. Unauthenticated clients cannot communicate with internal network entities because of address filtering (see Figure V.2, a). The purpose of this action is to separate the traffic of authorized and unauthorized users in order to protect the former.

In this case the access network is divided into two (or more) logical networks. The access process consists of the following phases:

1. Association with an access point;
2. Guest IP address acquisition;
3. PANA authentication;
4. Key establishment between the access point and the mobile node;
5. User IP address acquisition and
6. Updating address information at the PAA.

As the user can communicate with nodes in the internal network before being authenticated, many attack possibilities are open. Other shortcomings of the scenario are:

1. All "guest" network communications are insecure until cryptographic keys are negotiated between the AP and the MN;
2. Double address acquisition increases handover time and
3. The DHCP server is situated in a "demilitarized zone" (DMZ), all unauthenticated users have access to it, and the service is vulnerable to different kinds of attacks.

According to [136], the PAA and the authentication server, the PAA and the Enforcement Point (EP) have a priori trust relationships and it is natural to assume that paths between them are protected. An arriving PANA authentication Client (PaC) does not trust any network entity.

To reduce authentication latency and vulnerability of internal network entities, a modified architecture may be used (Figure V.2, b).

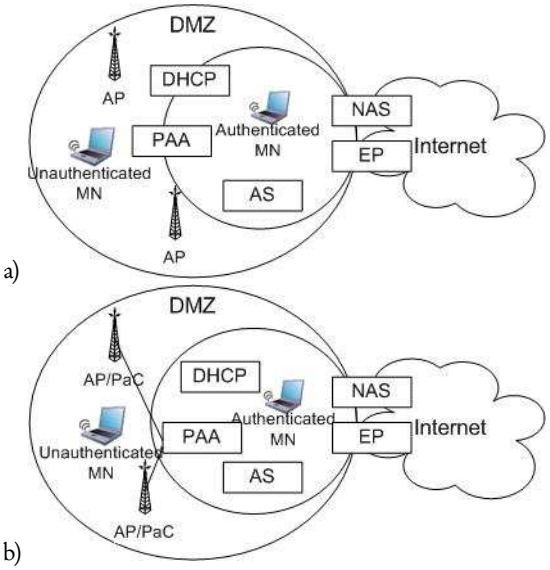


Figure V.2: Authentication infrastructure: a) PANA model, b) modified model

In the proposed architecture an access point has trust relationships with the PAA (via IPSec or TLS). All network entities having an IP address are in the protected internal network. Unauthenticated mobile nodes and all access points are situated in a kind of "quarantine zone". A non-authenticated mobile node has no IP address in the candidate network; it associates with an access point that opens a communication port only for authentication messages. The access point asks for the mobile node's identity and acts as a PaC, sending messages to the PAA. EAP authentication is executed between the mobile node and its home authentication server via a local authentication server, the PAA and the AP/PaC.

A combination of 802.11i and PANA protocols was chosen because the 802.1X standard provides a way to secure layer-2 encryption and integrity keys establishment between the mobile node and the access point. Non-authenticated mobile nodes must not have access to any network entity (see Figure V.2), and this is achieved by using the 802.1X controlled/uncontrolled port scheme. PANA transports authentication messages and grants or refuses network access to a user. It does not provide key establishment for layer 2 communications. PANA and 802.11i share out tasks: the former is employed for user authentication while the latter is for key negotiation and granting general network access. The authentication process is proposed for the first mobile node's authentication in an administrative domain, which is longer than subsequent ones in the same network.

There is much work to be done to develop a secure context transfer scheme between administrative domains [27, 137]. It is quite difficult to deliver any secure information from one access point to another in different domains, because access points often have only internal (non-routable) IP addresses and cannot be directly reached from an external world. Another problem concerns establishing secure communications between access points in different domains. That is why access routers are attractive candidates to actively participate in inter-domain context transfer and it is therefore desirable to place authenticator functionality at the access router.

For a proposed model the following assumptions have been made:

1. All resident entities in an “internal” network have trust relationships and strong security associations. Paths between the access point and the PAA, the PAA and the EP, the PAA (if they are not integrated) and the local authentication server must be protected by IPSec or TLS tunnels;
2. A visited network should have a certificate that is understood by a visitor;
3. There are roaming agreements between administrative domains, where a mobile user can nomad, so that when a mobile node presents its credentials, a local authentication server in a visited network can recognize a mobile node’s home authentication server and
4. A local authentication server puts authentication information into a cache for each visitor.

The second requirement is not too realistic, but if it is assumed that there are no a priori trust relationships between a mobile node and a visited domain, we must solve two tasks: establishing dynamic trust relations between domains, and visited network identity verification by the mobile node. This assumption allows one part of the mobility management problem described to be worked out.

V.1.3 Authentication process

The proposed authentication approach includes operations of IEEE 802.11i, PANA and RADIUS/Diameter protocols. Figure V.3 depicts a full authentication process using EAP-TLS method. This authentication method is set by default for Windows XP users, provides strong mutual authentication, higher performance than EAP-TTLS, and does not require a user’s interaction.

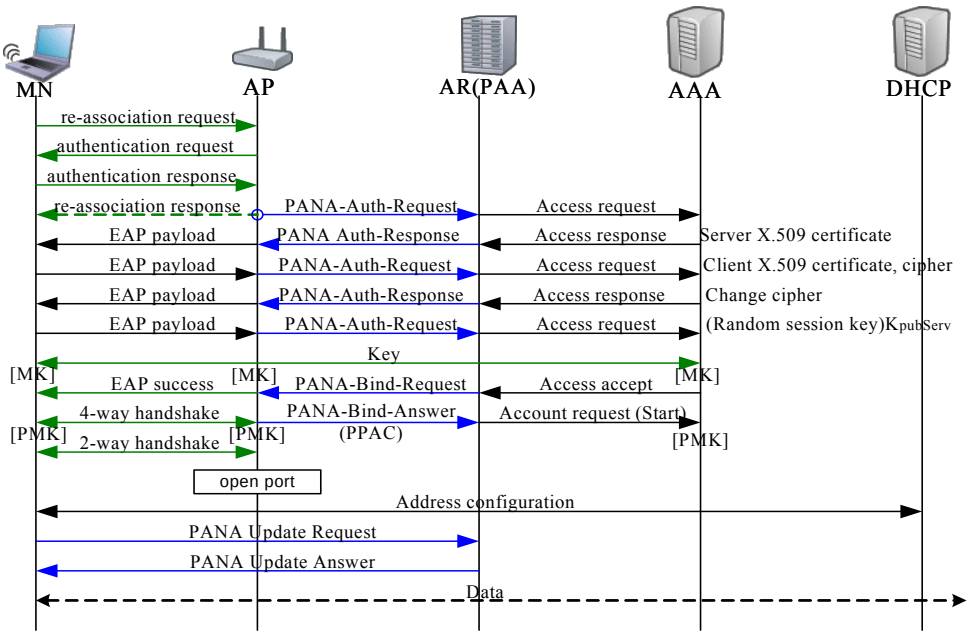


Figure V.3: Authentication exchange, EAP-TLS method

Several modifications are proposed to the initial methods. An access point, communicating with the mobile node, acts as an 802.1X authenticator, and, communicating with the PAA, acts as a PaC, sending PANA messages to the PAA, instead of RADIUS messages to a local authentication server. A discovery and

handshake phase is eliminated from the PANA message exchange, since the access point knows the PAA address and there is a secure channel between them.

The mobile node presents its identity in the form of the Network Access Identifier (NAI) [138], which helps a local AS to find a mobile node's home authentication server: *user@home_domain.com*. The paper does not concentrate on optimization of communications between the local authentication server and the mobile node's home authentication server.

After a re-association process, an access point connects to a PAA. The AP acts on behalf of the user terminal, and transmits a mobile node's device identifier to the PAA. The following authentication process is carried out in the usual way. In the PANA-Bind-Answer the Post-PANA address configuration option defined in PANA IETF draft [59] must be indicated to inform a PAA that a mobile node will change an IP address.

The PANA authentication process is optimized according to [122]: all PANA-Auth-Answer messages carry EAP payload instead of acting as an acknowledgement. This optimization is possible because there is a channel between the access point and PAA; communications are carried by wired media over a short distance, so there is a very low probability of packet loss.

Normally, the visited network is not a home for the mobile node, so a local authentication server must operate in proxy mode. This proxy authentication server may either know a path to a mobile node's home authentication server (if there are roaming agreements) or know a path to a central authentication server, which redirects it to the mobile node's home authentication server. Communication with the authentication server in a mobile node's home network significantly increases the overall authentication time on account of round-trip time that can be high value. Optimization of inter-authentication server communication and routing is outside the scope of this paper.

V.1.4 Performance analysis

PANA packet retransmission timer values are too large to meet fast handover requirements (the Initial Retransmission timeout is 1 sec, Maximum Retransmission Timeout is 30 sec [137]), taking into account the high probability of packet loss in a wireless network. If traffic is managed by an access point at the MAC layer, detection of lost packets and their retransmission takes less time (the minimum value of acknowledgement timeout is about 3 ms, the maximum value is about 52 ms). We assume that the probability of packet loss is much smaller on the wired connection between the access point and the PANA Authentication Agent then on the wireless connection between the mobile node and the Access Point.

In this evaluation we consider three implementation scenarios:

1. Address filtering without link-layer authentication,
2. Consequent link-layer and network-layer authentication and
3. The proposed compound authentication.

The number of messages exchanged between the mobile node and the Authentication Server and sent via the authenticator depends on the EAP method executed. Independently of the implementation scenario the authentication latency is

determined by the latency of message transmission between the mobile node and the authenticator (T_{MN_Auth}), the latency of message transmission between the authenticator and the Authentication Server (T_{auth_AS}) and time of message processing by the mobile node, the authenticator and the Authentication Server (T_{proc_MN}, T_{proc_AP}, T_{proc_AP}, T_{proc_AS}). Figure V.4 depicts operations and corresponding time latencies for these scenarios. T_{MN_AP}, T_{MN_PAA}, T_{AP_PAA}, T_{PAA_AS}, T_{AP_AS} are times to transmit a message between the mobile node and the access point, the mobile node and the PANA Authentication Agent, the Access Point and the PANA Authentication Agent, the PANA Authentication Agent and the Authentication Server and the Access Point and Authentication Server respectively.

For the first scenario the authenticator is the PANA Authentication Agent and the transmission latency T_{MN_Auth} for each message is defined by T_{MN_PAA}. In the second scenario, this time is T_{MN_AP} and T_{MN_PAA} correspondingly for each step and in the proposed approach this time is composed by T_{MN_AP} and T_{AP_PAA}. Each time value represents an amount of time spent by the specified entity during the all authentication process. Thus, T_{MN_AP} means the sum of latency of messages transmission between the mobile node and the Access Point during authentication.

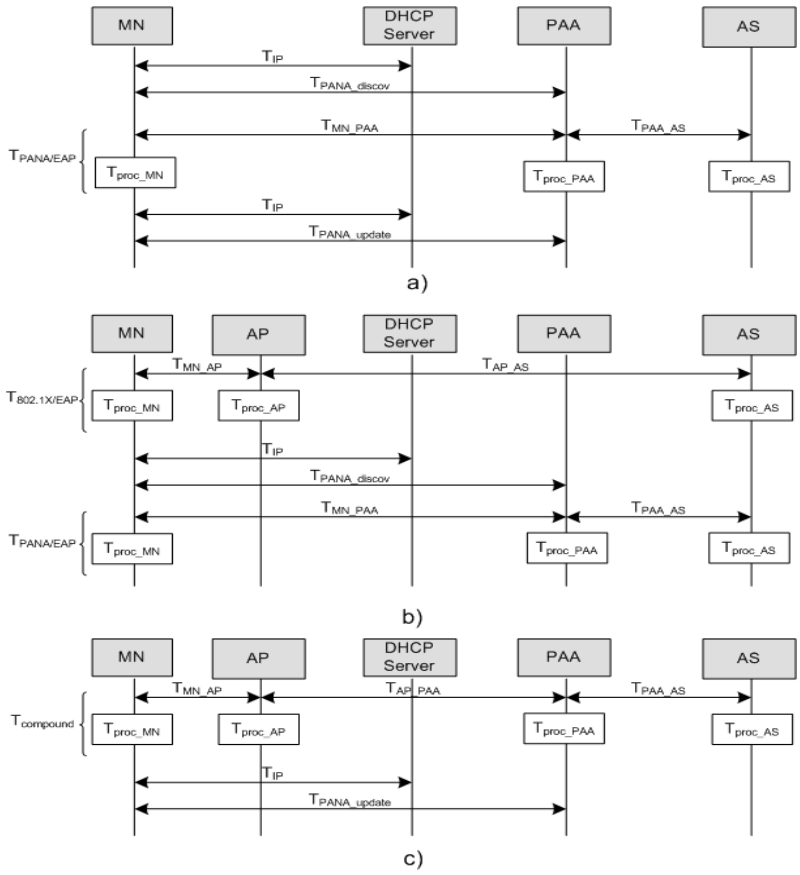


Figure V.4: Authentication latency for scenarios with address filtering (a), consequent link-layer and network-layer authentication (b) and compound authentication (c)

As all entities execute the same methods in different scenarios, the message processing time and message exchange time are supposed to be equal for all cases. Links are supposed to be symmetric.

As compared with PANA authentication, the proposed authentication gains the time taken by the PAA Discovery phase and loses the time taken by access point message processing, which is relatively small.

In comparison with the first scenario (Figure V.4, a) the proposed approach (Figure V.4, c) allows avoiding double IP address acquisition and eliminates PANA discovery phase. On the contrary, compound authentication takes longer time:

$$T_{compound} - T_{PANA/EAP} = T_{proc\_AP} + T_{MN\_AP} + T_{AP\_PAA} - T_{MN\_PAA}, \quad T_{compound} - T_{PANA/EAP} > 0 \tag{V.1}$$

As against the second scenario (Figure V.4, b) compound authentication demonstrates latency reducing:

$$T_{compound} - (T_{802.1X/EAP} + T_{PANA/EAP}) = T_{AP\_PAA} - (T_{proc\_MN} + T_{proc\_AS} + T_{MN\_PAA} + T_{AP\_AS}), \quad T_{compound} - (T_{802.1X/EAP} + T_{PANA/EAP}) < 0 \tag{V.2}$$

A situation where the authenticator communicates with a remote Authentication Server is also possible. From (Eq.V.1 and Eq.V.2) it can be seen that the time of message transmission between the authenticator and the Authentication Server does not change the ratio of authentication latency in analysed approaches.

In order to estimate the time taken by the proposed authentication approach and to compare it with the performance of approaches introduced earlier, we analysed EAP-TLS method execution. The parameters of authentication model were defined as follows: the processing time for the 802.1X authentication is 150 ms, for PANA 160 ms. For the DHCP operation the assumed time is 200 ms. Figure V.5 depicts the authentication latency in case of address filtering, consequent link-layer and network-layer authentication and the compound authentication. The time taken by both scenarios is shown in Figure V.5.

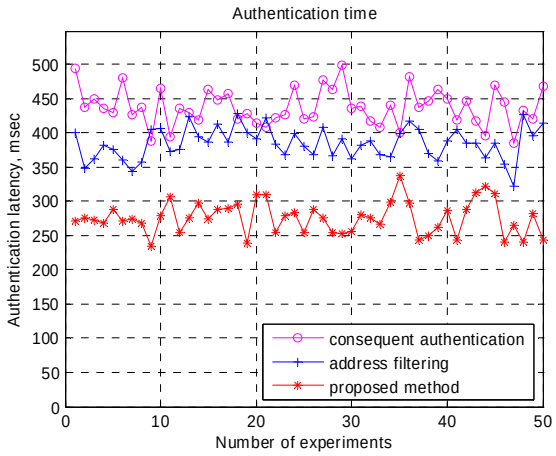


Figure V.5: Authentication time for authentication with address filtering, consequent and compound link-layer and network-layer authentication

The proposed approach is more efficient then previous proposals because it ensures link-layer security of communication and reduces the authentication latency compared to the consequent authentication scenario. The number of messages

exchanged and processed is the same that in the case of PANA use. The compound authentication introduces additional processing time to the Access Point operation.

V.1.5 Summary

Parallel authentication permits a mobile user to obtain Internet access as a result of a single authentication in a multi-service network. The proposed approach combines the operation of the two most commonly used protocols to authenticate a user to a network and a service, and provide strong link-layer encryption for communications. An access router is a good candidate for the role of authenticator because this scheme may serve for pre-authentication using context transfer between different administrative domains.

The proposed approach does not allow communication between an unauthenticated mobile node and internal network entities. It aims to protect the DHCP server and access router from untraceable DoS attacks. The performance of the process may be improved due to the exclusion of a PAA discovery and handshake phase from authentication and double IP address acquisition. The security level is not compromised; all communications inside the network are secured.

We do not take into account a time interval taken in searching for and communicating with a mobile node's home authentication server, as it concentrates on local authentication and improvement of security of network access.

V.2 FAST RE-AUTHENTICATION PROTOCOL: A SOLUTION FOR INTER-DOMAIN AUTHENTICATION

The primary goal of our work is to minimize user authentication time in a visited network. The proposed method is based on symmetrical cryptography and uses challenge-response mechanism. FAP includes identity privacy support and traffic encryption key generation. The access to the target network is granted to the user if the latter proves that he has been recently successfully authenticated with the roaming partner of the target network. This proof is contained in the ticket, which is given to the user by the partner of the target network. In the encrypted part the ticket contains information known to the user, and only the issuer and the addressee can decrypt this information. In such a way the user is able to check the identity of the target network. FAP should also establish secure connection between two authenticated parties.

We specify an Extensible Authentication Protocol (EAP) mechanism for fast authentication in inter-domain roaming. FAP is technology independent and may be implemented over any wireless network (802.11 802.16 or 3GPP). We merely illustrate the protocol operation in 802.11 networks.

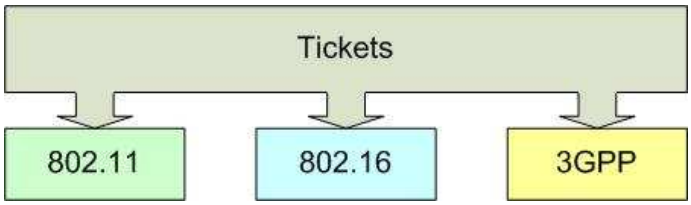


Figure V.6 : Technology independence of the ticket

The proposed protocol consists of two sub-protocols: ticket acquisition and fast re-authentication. The former is executed when the user is attached to the network and it requires inter-domain communication, and the latter is executed during handover and localizes the authentication process in the target domain

V.2.1 Assumptions apply

The mobile user can roam from one non-home network to another. To distinguish these visited networks we will call the one where the user has been authenticated the current network and that which the user is roaming the target network. We also assume that there are roaming agreements either between the home and the target networks or between the current and the target networks. The user can communicate in a secure manner with his home domain. Authorities that have roaming agreements share symmetric or asymmetric keys $\{K_{Rij}\}$. In the situation where brokers are used, the user may solicit authentication tickets from a known broker and this procedure is equivalent to the communication with the home network.

The operation of FAP is based on the assumption, that the mobile node is attached to a network and has performed an initial, full authentication by some other means. The protocol is focused on fast re-authentication during inter-domain handover.

The current network of attachment may generate tickets for a client as well as its home network. We call the entity, which is able to distribute tickets to a particular user, the anchor FAPS and, to simplify an explanation, denote it as FAPS.

The operation of the proposed protocol does not depend on the nature of the security associations between partner domains. Thus the authentication ticket may be encrypted with both a symmetric an asymmetric key.

The authentication protocol uses digital signing and block-cipher (CCM [139]) encryption cryptographic operations. In digital signing, one-way hash function HMAC-SHA-256 [140] is used.

V.2.2 Roaming scenarios

In a roaming scenario the following situations are possible:

1. The mobile node is attached to its home network. In this case the home FAPS will generate tickets for all its neighbouring partners.
2. The mobile node is attached to a network that is a partner for its home network. After successful authentication the client requests tickets from its home network. When the mobile node is attached to a visited network, it may communicate with its home network for one of the following reasons: authentication or location update.

We assume that the mobile node and its home network share some secret information. This information can be the result of authentication or a pre-shared key.

If the mobile node is originally attached to the visited network, it has performed successful authentication in this network and both the user and the network have generated key material.

The user receives tickets from the visited network for its neighbours only. The current network has a responsibility to decide whether tickets for its partners should be sent to the authenticated client. This decision is based on the nature and rules of roaming agreements between the current, target and home domains.

The mobile node is attached to a partner of the previous visited network, which is not a partner for its home network.

From the point of view of the roaming destination, the mobile node may choose a network, which is

- ✓ A roaming partner for its home operator;
- ✓ A roaming partner for its current visited network;
- ✓ A roaming partner for its home and its current visited networks.
- ✓ Neither a partner for the home nor for the current network.

V.2.3 Architecture overview

Fast re-Authentication Protocol specifies communication between the FAP Server (FAPS) at the network side and the FAP Client (FAPC) at the user's side. The generalized architecture of FAP is shown in Figure V.7.

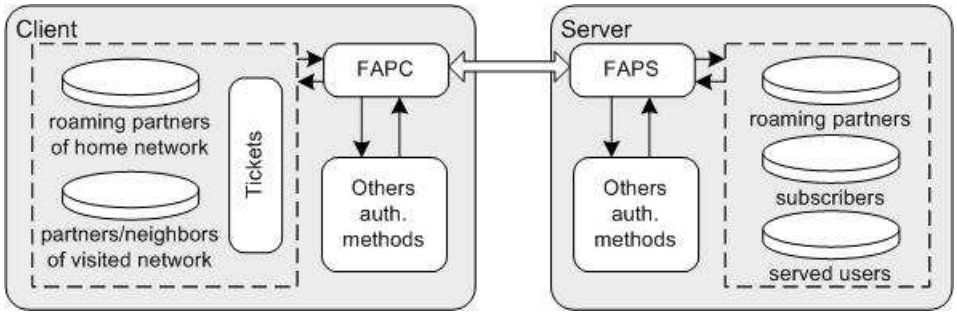


Figure V.7: Architecture of the client and server parties of FAP

On the client's side. The client's part of the protocol has access to the information produced by other authentication methods, to the database of roaming partners, and to the database of neighbours of the visited network. The information from other authentication methods are used for derivation of credentials for fast re-authentication. Knowledge of roaming partners of the home network and partners of the currently visited network helps the mobile node to choose the target network among candidates. Information about partners of the current network changes after each authentication in a visited domain, while the information about partners of the home network changes only when a contract with an authority is signed or terminated. The procedure of the update of this information is outside the scope of this work.

On the server's side, FAPS communicates with databases of roaming partners and the database of subscribers of its network. It also communicates with the list of currently served users.

The database of roaming partners for each partner contains its name, the list of its name as displayed to clients, parameters of the corresponding security association and roaming policies. This information is kept for a long period of time. The FAPS communicates with this database on receiving an authentication request (Ticket message).

Each record of the database of subscribers contains the subscriber's identity, the subscriber's roaming pseudonym, the shared secret, information about current location. The latter includes the data related to the last authentication. The FAPS communicates with this database on receiving a Ticket request from its subscriber.

A user being served may be a subscriber, a visitor, for whom the FAPS may generate a ticket, or a visitor who is not authorized to receive a ticket according to the rules of the roaming agreements with his home network.

To generate authentication tickets, the FAPS should have access to results of different authentication methods, which may have been used for the last authentication.

V.2.4 Ticket acquisition

The ticket acquisition phase is designed to provide a user with credentials for further fast authentication. The network generates tickets after user authentication and does not care about ticket renewing and revocation. If the ticket expires, the new one is generated exclusively upon the user's request. The user can roam either to a partner of his home network or to a partner of the current network. The current network (home or visited) may generate tickets for the user.

Authentication ticket format. The idea of the method is to use a short-lived, lightweight ticket, which does not require a revocation mechanism and may only be verified by the issuer and the target network. The authentication ticket is proposed to decrease handover latency, which is why its calculation and verification procedures should not be computationally heavy. The ticket format is presented in Figure V.8.

C: part in-clear	
target_name	72 bytes
issuer_name	72 bytes
expires	6 bytes
S: encrypted part {	
auth_res	32 bytes
user_pseudonym	72 bytes*
}K_R	
	254 bytes
Signature SHA-1(C|S, K_R)	***32 bytes***

Figure V.8: Ticket format

The ticket is bound to the issuing and target networks by usage of the key K_{Rij} shared between two domains *i* and *j*. It is also bound to the user by user-pseudonym and the previous authentication result, which are described further.

The ticket consists of two parts. The section S (hereinafter called secret) is encrypted with the key K_R that is shared with a particular roaming partner of the ticket issuer. The authentication result "*auth_res*" is produced from information related to the previous authentication as shown in (V.5). The maximum length of this field is 256 bits (32 bytes). As the target FAPS (tFAPS) must obtain the user name [141], the latter is presented in the ticket. On the other hand, the identity of the user should be hidden. To satisfy this requirement the "*user_pseudonym*" is a roaming pseudonym of the user. This pseudonym is the user identity perceived by the visited network in the initial authentication. It is not equivalent to username in general cases. Nevertheless,

the user-pseudonym must contain the link to its home network. This information is used for accounting purposes and based on it the visited network makes a decision if it will create tickets for this client.

Part C (see Figure V.8) is not encrypted. It contains "*target_name*" that is the name of the network, which is able to decrypt this ticket, "*issuer_name*" that is the name of the network, which has provided the ticket and the "expires" field, which determines the end of the ticket validity period in the form {day month year hours minutes seconds}. This ticket expires after a short period of time (defined by the issuer) and cannot be renewed automatically but only upon user request.

The "*target_name*" represents the identity name of the partner of FAPS. As one authority can manage several networks (i.e., UMTS and WiFi hotspots) its name display may vary on different interfaces. To make the ticket format technology independent (See Figure V.6) and to avoid generation of more then one ticket for each roaming partner, the FAPS provides the FAPC with a function that matches different seen names of a network with the "*target_name*":

$$\{seen\ name\ 1, \ldots seen\ name\ k\} \rightarrow target\_name \qquad \text{(V.3)}$$

The cFAPS knows which "*seen_name*" is visible and sends the FAPC the correspondence between these names and the "*target_name*" contained in the ticket. The FAPC may permanently hold the function provided by its home FAPS and the latter does not care about the nature of the current neighbourhood of the subscriber.

The entire ticket is signed with the key K_{Rij} to assure its integrity protection. For encryption and signature the FAPS may use either a single key or separate keys against the security association between the partner networks.

Figure V.9 depicts principal steps of FAP operation and information exchange. When the user terminal is attached to a network, we assume that strong mutual authentication is completed between them (it may be either the initial authentication or re-authentication after FAP accomplishment). In this situation the user terminal trusts its home domain via certain shared data and the current domain via the result of the recent authentication. The current FAPS generates or not tickets for guest users according to the presence of agreements with the user's home FAPS and policies of the current FAPS. The user terminal sends a Ticket request message to the home network and to the current network. In a case where the latter does not support ticket generation for guests it responds with a Ticket reject message. We suppose that the FAPS always generates tickets for its operator's subscribers.

In the given example we demonstrate ticket generation only for one partner (tFAPS) of the MN's current network of attachment (cFAPS). Upon a single Ticket request the cFAPS generates tickets for all its neighbouring partners. Considered FAP servers share strong key K_{Rct}.

Mobile node keeps some *method_res*, the data related to the last authentication, also kept by the authenticated network. The server, trusted by the user, creates an authentication ticket, which contains the result of the previous authentication *auth_res*.

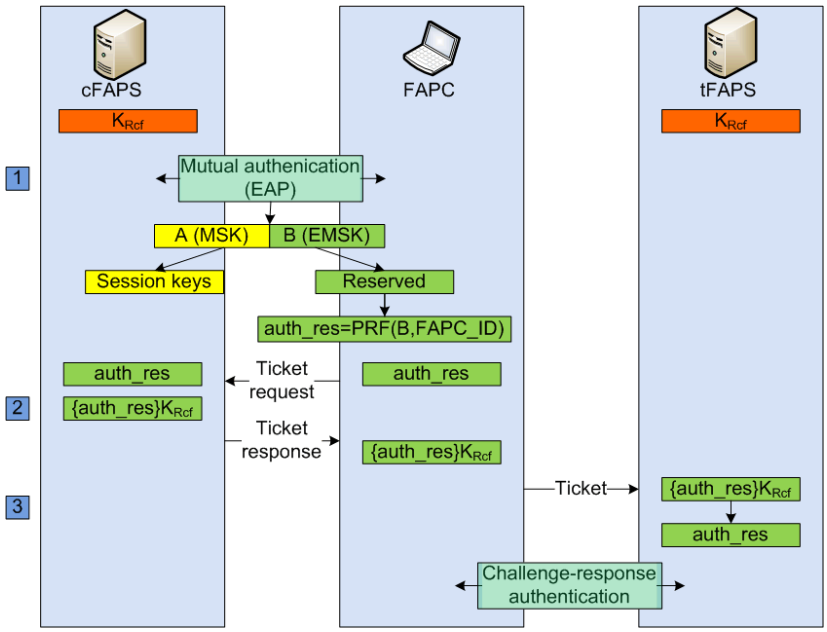


Figure V.9: FAP operation sequence

The *auth_res* is derived from the *method_res* both by either the FAPS and by the FAPC as (Eq.V.4) shows. The symbol "||" denotes concatenation. The pseudo-random function (PRF) is calculated according to the algorithm described in RFC 2104 [142].

$$auth\_res = PRF(method\_res, user\_pseudonym \,||\, MN\_MAC) \qquad (V.4)$$

The *method_res* contains information negotiated as a result of the MN's authentication in a previous network. If EAP authentication is executed, two keys are produced: Master Session Key (MSK) and Extended Master Session Key (EMSK). The latter is not used in session keys derivation and that is why it can be used for fast re-authentication purposes.

The FAPS encrypts *auth_res* and *user_pseudonym* with a key K_{Rcf}, shared with a particular roaming partner. It completes the ticket with the date and the time of ticket expiration, target domain name and its own domain. Finally the FAPS signs the entire ticket with the same key K_{Rcf} and sends it to the FAPC upon Ticket Request. The FAPC is not able to decrypt the secret part of the received ticket.

Each FAP server keeps a list of roaming partners and a list of subscribers that change only when a subscriber or a roaming partner is added or eliminated. After each successful authentication a FAPC keeps two lists of roaming partners of the hFAPS and the cFAPS, which are reachable from the current user's location. Each list contains corresponding tickets. This information is updated when a user is authenticated in a new network.

The visited network has a responsibility to decide whether the list of roaming partners with corresponding secrets must be sent to the authenticated client. This decision is based on the nature and rules of roaming agreements between the current, the target and the home domains.

The knowledge of the neighbourhood of the client's current network of attachment of the client may be used to reduce the number of tickets generated and sent to each

user. We call two networks neighbouring if users can handover from one network to another. If the network knows the current location of its subscriber and it knows which partners adjoin with this network, it generates and sends to the user tickets only for these partners.

V.2.5 Re-authentication protocol

The authentication protocol provides authentication of the client and the visited network without communication between the target and the user's home network. The protocol provides secure negotiation of a shared secret. The attacker cannot modify the communication without being detected by the parties.

Before starting the authentication process the client (FAPC) possesses an encrypted and signed ticket, which may be verified by the target network. To find a ticket corresponding to a selected target network the client uses the function matching the perceived network name with the name of the managing authority. Figure V.10 gives an example of ticket selection for a target network. In the given example the client may handover to two networks. UMTS and 802.11 managed by the same operator. Due to the technology independence of a ticket the same credentials are applied to both access networks.

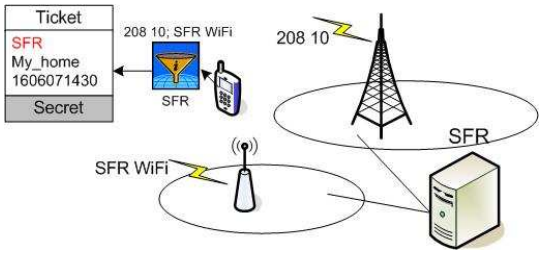


Figure V.10: Choosing a ticket

The target server (tFAPS) has a key to decrypt and verify the ticket. Figure V.11 shows the information flow in the authentication exchange.

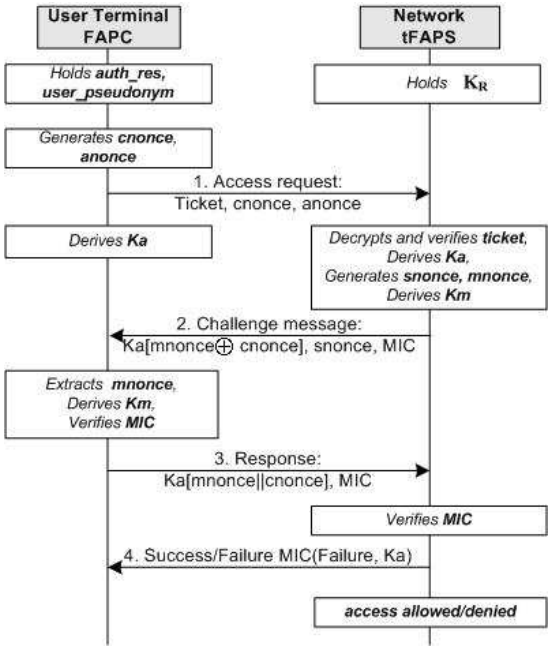


Figure V.11: Flow chart of the FAP authentication exchange.

Assumptions and cryptographic functions

cnonce and *anonce* are random numbers generated by the FAPC.

snonce and *mnonce* are random numbers generated by the FAPS.

K_a is the authentication key, which is derived from the information contained in the ticket *auth_res*, the random number *anonce*, the address of the mobile node's network interface *cAddr* and the *user_pseudonym* according to the algorithm described in RFC 2104 [142]:

$$K_a = PRF(auth\_res, "authentication\quad key", anonce \,||\, MN\_Addr \,||\, user\_pseudonym) \tag{V.5}$$

K_m is the Master Secret, which is generated in case of successful authentication and serves as a material to session keys derivation. This key is calculated as follows (Eq.V.6):

$$K_m = PRF(auth\_res, "master\quad \sec ret", \min(anonce, mnonce) \,||\, \max(anonce, mnonce) \,||\, user\_pseudonym) \tag{V.6}$$

The *MIC* denotes Message Integrity Code; it is computed over the body of the message (denoted as *msg*) using the Master Secret K_m as shown in (Eq.V.7).

$$MIC = HMAC - SHA - 256(K_m, msg) \tag{V.7}$$

Message exchange

The FAPC sends Authentication request **message** to start authentication process with the tFAPS, after the ticket, corresponding to the target network, is found. This message contains user credentials and provides the tFAPS with the material for further key generation. After sending the Authentication request message containing a ticket, the FAPC calculates an authentication key K_a. On reception of Ticket

message the tFAPS searches in its database of roaming partners a key shared with a domain mentioned in the Ticket message. If the domain name is found, it decrypts the ticket with a corresponding key K_R and calculates K_a in the same way as a client. The tFAPS generates a random value *snonce* and derives a Master secret K_m, as shown in (Eq.V.6).

The tFAPS cancels authentication and responds with a Failure message if the mentioned authority is unknown, if tFAPS cannot decrypt the ticket or if the ticket has expired.

The tFAPS responds with **Challenge message** to the FAPC. This message contains the result of XOR function of *cnonce* and *mnonce*, encrypted with K_a, the *snonce* and the integrity code of the entire message, are computed using K_m according to (5). Sending this message, the tFAPS proves its identity to the FAPC. On reception of this message the FAPC extracts the *mnonce*, derives K_m in the same way as the tFAPS and verifies the message integrity code. If the computed and received values of MIC do not match, verification fails. That is possible if K_a is not derived correctly, if *cnonce*, used by the tFAPS, is not valid or K_m is not derived correctly. In this case the FAPC sends the Failure message containing the error indication.

If verification was successful, the FAPC sends a **Response message** to the tFAPS. This message demonstrates to the tFAPS that the client is live and allows the tFAPS to verify if the FAPC has derived the same Master secret K_m. The tFAPS responds with a **Success message**, if the calculated MIC matches the MIC included in the Response message. Otherwise the tFAPS sends **Failure message** to the FAPC.

The Master secret K_m may be used for further generation of session keys.

If the target network does not support FAP, the mobile node should perform authentication using a method supported by the network.

V.2.6 Implementation of the fast re-authentication protocol

To illustrate FAP operation we have chosen IEEE 802.11 technology. Since most authenticators support the 802.1X [51], it is natural to build the authentication phase of FAP on top of the 802.1X framework to avoid modifications on the authenticator's side. The implementation of our approach is based on the introduction of a new EAP method, called EAP-FAP. To support this method, modifications have been made at the supplicant and at AAA server.

To minimize the number of messages exchanged, we have extended the *EAP Response Identity message*. The field containing the identity string is extended by zero byte and the FAP authentication ticket. According to Section 5.1 of RFC 3748 [52], this message may contain additional options and should not be larger than 1020 octets. In case of an unknown identity or invalid authentication data, the authenticator communicates the reason for the failure in the EAP-Nak message to the supplicant.

The general packet format of EAP-FAP is shown in Figure V.12. The white fields represent standard EAP fields [52], and gray fields represent the FAP header and FAP Data, which are contained in the Type-Data field of the EAP packet.

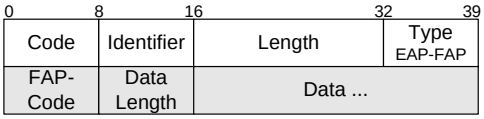


Figure V.12: EAP-FAP packet format

The *Code* field contains *1* for EAP Requests and *2* for EAP Responses.

The content of the *Identifier* field is identical to any other EAP method.

The *Type* field should be set to the assigned value for EAP-FAP.

The *FAP-Code* field may take on values of *1* (Challenge) and *2* (Response).

The size of the *Data Length* field is one octet as the maximum length of *Data* is less than 256 octets.

The *Data* field contains either a challenge or a response for it according to the *FAP-Code*.

If either the candidate network's authentication server cannot decrypt the ticket or the function value in the third message differs from the estimated value it responds with a *Failure message* containing a *Reason code*. Possible reason codes are listed in Table V.1.

Table V.1: Reason Codes meaning

Reason Code	Meaning	Message in a sequence
0	Unknown authority	2
1	Invalid authentication data	2
2	Verification failure	4

We have set up our test-bed to estimate the delay of the authentication phase of FAP and to compare its performance with TTLS-MD5 authentication protocol.

V.2.7 Experiment results for FAP implementation

V.2.7.1 Test-bed setup

In our test-bed, we used a RADIUS server that works under FreeRadius [143] software. For supplicant implementation, we have chosen Xsupplicant [144], which is open source 802.1X client realization. We modified this software by adding a new EAP method, called EAP-FAP. We have implemented the user part (EAP-FAPC) and the server part (FAPS) of the proposed authentication method. The authenticator software was not changed. Figure V.13 shows components we used in the protocol implementation.

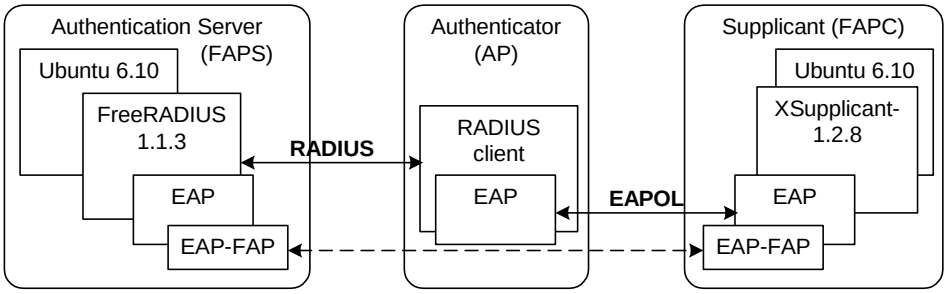


Figure V.13: Scheme of FAP implementation at each network entity participating in authentication

Authentication server – RADIUS server

hardware

Intel Pentium(R) 4 CPU 1.5GHz, 256Mb RAM

software

Linux Ubuntu 6.10
FreeRadius-1.1.3

Authenticator – Access point

hardware

CISCO AIRONET 1200 series, supports both 802.1X and RADIUS

Supplicant

hardware

Intel Pentium Centrino , 512Mb RAM
Intel(R) Pro/Wireless 2200 BG Network Connection

software

Linux Ubuntu 6.10
Xsupplicant – 1.2.8

V.2.7.2 Implementation Details

Server-side: FreeRADIUS

Configuration: EAP-FAP is implemented as module under FreeRADIUS-1.1.3. This method is referenced in the configuration file

`/usr/local/etc/raddb/eap.conf`

as a default eap type:

```
default_eap_type = fap
fap {
}
```

Code description: All modifications and additions were made in the `src/modules/rlm_eap` directory of FreeRADIUS source package.

We have defined a new type named FAP in the file `libeap/eap_types.h` and integrated it in the file `libeap/eapcommon.c`. The server part of the FAP is defined in the new directory `/types/rlm_eap_fap`. `eap_fap.c` determines FAP functionality and `rlm_eap_fap.c` lists handlers called by EAP module.

Client-side: Xsupplicant

Configuration: before starting the modified version of xsupplicant, it must have a correct configuration file.

`xsupplicant.conf`

```
identity = maryna

eap-fap {

username = maryna

ticket = /etc/tickets a   path   to   a   directory   containing
authentication tickets

auth_res = /etc/auth_res     a path to a directory containing the
result of the previous authentication

}
```

Code description: we have modified the source code of xsupplicant-1.2.8. The directory `/types/fap/` contains `eap-fap.c` file, which defines the functionality of the client part of `eap-fap`. We have extended the file `eap.c` in order to add the ticket to the EAP Response Identity message. Files `xsupconfig.h`, `config_lexicon.l` and `config_grammar.y` were modified in terms of variables, types, tokens and structures definitions.

V.2.7.3 Experiment results

We set up our test-bed to estimate the delay in the authentication phase of FAP. We implemented EAP-TTLS with MD5 in the second phase, MD5 and the proposed protocol (FAP), and we measured delays for 100 authentications resulting in an average latency of 85.33 ms for TTLS, 20.72 ms for MD5 and 30.59 ms for FAP. Figure V.14 shows authentication latencies observed over time for the protocols studied. Local latency maximums are caused by other applications run at the host and network load on the same interface. Authentication latencies were measured by capturing packets using the WireShark [145] network analyzer.

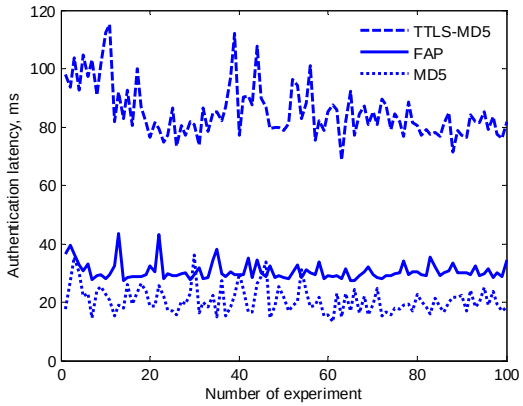


Figure V.14: Authentication latency for FAP, TTLS with MD5 and MD5

We have evaluated the authentication phase of FAP, and therefore did not take into consideration the time of association to the access point and the time of key

negotiation (as the corresponding algorithm was not be modified). The obtained authentication latency represents the time elapsed between receiving the EAP Request Identity and EAP-Success by the supplicant. The access point includes information about supported authentication methods (802.1X or WEP) according to [14]. The supplicant sends a modified EAP Response Identity on the authenticator's EAP-Request Identity. If the target authentication server supports FAP, it continues the authentication, otherwise it responds with a NAK message and the supplicant has to perform a full authentication using the supported method.

V.3 OPTIMAL CREDENTIALS DISTRIBUTION FOR INTER-DOMAIN AUTHENTICATION

To minimize the number of secrets exchanged between the mobile node and its home network we introduce a ticket distribution scheme for Fast re-Authentication protocol (FAP) for inter-domain roaming. FAP is designed to reduce the authentication time of a mobile user in a visited administrative domain. The approach eliminates the need for communication between the visited network and the subscriber's home network for credentials verification and uses a short-living lightweight re-authentication ticket, which does not require a revocation mechanism.

To minimize the number of authentication tickets sent to each subscriber, we propose the use of a neighbour table, which is maintained by an authentication server of each network. When the client requests a ticket, the server generates tickets only for the networks contained in the line of the neighbour table corresponding to the current location of the user. This method decreases the number of tickets sent and, consequently, the overhead and the delay of the ticket acquisition phase of the protocol.

To create this neighbour table, we propose a reactive mode for the ticket acquisition phase. In this mode, the mobile node chooses the target network for which it does not have a correspondent ticket. Then the mobile node sends a Ticket request to its home network indicating the current location and the chosen target network's name. The FAP server responds with a ticket for indicated target network and adds the name of this network to the list of neighbours of the user's current network of attachment. While the ticket acquisition protocol operates in the proactive mode, the mobile node sends the Ticket request without indicating the target network. The hFAPS sends tickets for all known neighbours of the current network of attachment of the mobile node.

V.3.1 Neighbour table construction

The home network creates secrets for all of its roaming partners. To reduce the number of tickets sent to the user after each authentication, the home FAPS (hFAPS) may keep the list of neighbours for each roaming partner. Networks are referred to as "neighbouring networks" if their coverage areas overlap and users can handover between them.

To minimize the number of authentication tickets sent to each subscriber, the hFAPS maintains a table of neighbours for each roaming partner. Each line in this table

contains names of roaming partners of the home network. When the FAPC requests a ticket, the hFAPS generates tickets only for those networks contained in the line of the neighbour table corresponding to the current location of the user. This approach reduces the number of tickets sent and consequently the overhead and delay of the first phase of the protocol.

Figure V.15 shows an example of the location of networks and the corresponding neighbour table. The line in the figure indicates the presence of a physical path from one network to another.

Before the neighbour table is created, the protocol operates in a reactive mode. The hFAPS sends tickets at the demand of the FAPC for the chosen target network, only if the latter is a roaming partner for the home network.

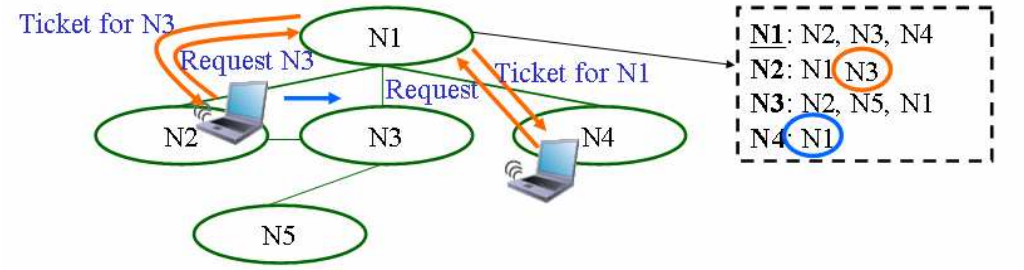


Figure V.15: Network neighbouring

Each FAPC keeps a list of roaming partners of its home network and after successful authentication in a visited network it has a (possibly empty) set of tickets from hFAPS and cFAPS. In the most optimistic scenario, the mobile node has not begun the handover procedure and it receives advertisements from the network, which is in the list of the home network's roaming partners, but the FAPC has no ticket for this network. In this case, the FAPC sends a ticket request to the hFAPS. If the roaming agreement exists, the latter responds with the ticket and adds the target network in the neighbour table. If the roaming agreement does not exist, the hFAPS responds with the corresponding error code, and the FAPC deletes the name of the network from the list of the home network's roaming partners.

In the less optimistic scenario, the mobile node begins handover and realizes that it has no credentials for fast authentication in the target network. The FAPC then executes the same procedure described in the previous scenario.

In the initial phase of the neighbour table construction, the user authentication process consists of ticket acquisition and authentication.

V.3.2 Formal validation of the model

In this section we present a formal performance analysis of reactive and proactive modes of FAP operation.

Let the roaming region be covered by n networks $\{N_i\} = (N_1, \ldots N_n)$. Table V.2 shows notations used in this section. Let us choose network N_i for further analysis and for simplicity's sake denote it N.

Table V.2: Used notations

Notation	Meaning
ns	Number of subscribers
nc	Number of clients served by the network
np	Number of roaming partners
$\{R_{ij}\}=(R_{i1},\ldots R_{ipi})$, $\{R_{ij}\}\subset\{N_i\}$	Set of roaming partners for the network N
v	Number of neighbours
$\{V_{jk}\}=(V_{j1},\ldots V_{jv})$	Set of neighbours of N_j
vp_j	Number of neighbours of the network N_j, which are partners for N
m	Number of elements in the network table
t_r	The average time of user residence in a network
t_{proc_auth}	The time of an authentication request processing
t_{proc_tick}	The time of a ticket request processing

V.3.2.1 Reactive mode

Before the neighbour table is created, the protocol operates in a reactive mode. The hFAPS sends tickets at the demand of the FAPC and only for chosen target network N_j, if the latter is a roaming partner for the home network $N_j \in \{R_{jp}\}$. In the proactive mode for j^{th} network, the hFAPS creates vp_j tickets.

$$vp_j=\deg(\{V_i\}\cap\{R_{jp}\});\ vp_j\leq np \tag{V.8}$$

As can be seen from Figure V.15, the finished neighbour table contains m elements:

$$m=\sum_{j=1}^{np} vp_j \tag{V.9}$$

To make the proposed authentication method efficient, the reactive mode of ticket acquisition should not take a long time. Users execute handovers between networks operated by roaming partners of their home providers. Each user chooses a target network among neighbours of the serving network with the uniform probability. Let us represent the process of the neighbour table creation as a chain of states, where each state corresponds to the specific number of partners added to the table. The system can change the state when a subscriber sends the reactive ticket request. Initially the neighbour table at the hFAPS contains only a column of roaming partners, and our system is in the zero state. After a user's ticket request, the hFAPS adds the name of the target network in the line corresponding to the network attached to the user, and the name of the current point of attachment to the line corresponding to the target network.

The following equation shows the probability of adding a new record to the neighbour table at any moment k.

$$P_i(k)=\sum_{j=1}^{m/2} P_j(k-1)\cdot\left(1-\frac{2\cdot k}{m}\right) \tag{V.10}$$

where ***m*** is the general number of records in the partner table, as defined in (Eq. V.9). We can interpret (Eq. V.10) as the probability of receiving a reactive ticket request at any moment.

V.3.2.2 Proactive mode

In addition to authentication latency, the performance of the proposed method is determined by the load of authentication servers.

We represent functionality of each authentication server as shown in Figure V.16.

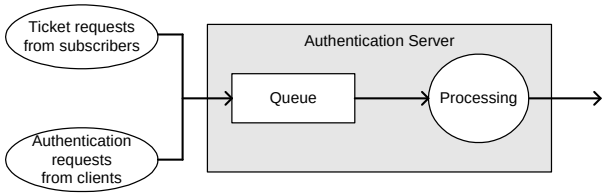


Figure V.16: Functionality of the authentication server

The server receives two types of requests: ticket requests from its subscribers and authentication requests from clients. Clients may be both its own subscribers and subscribers of its roaming partners. The maximum number of clients ***nc*** for network *N* is

$$nc=\sum_{j=1}^{np} ns_j \tag{V.11}$$

The operation of the authentication server represents a discrete-time stochastic process with the Markov property. In this process, each state of the system corresponds to the probability of a certain number of requests waiting in the queue. The server receives a request of any type with a frequency

$$\lambda=\frac{1}{np}\cdot\frac{nc}{t_r}+\frac{ns}{t_r}=\frac{nc+ns\cdot np}{np\cdot t_r}. \tag{V.12}$$

The flow of processed requests is defined as

$$\mu=\frac{t_{proc\_auth}+t_{proc\_tick}}{t_{proc\_auth}\cdot t_{proc\_tick}}. \tag{V.13}$$

When the neighbour table is created, the system works in the stationary mode, and probabilities of all states are time-independent. Reasoning from the values of request processing obtained from experiments

$$\rho=\frac{\lambda}{\mu}\leq 1. \tag{V.14}$$

Thus, the probability of ***i*** requests waiting in the queue is

$$P_i=\rho^i(1-\rho) \tag{V.15}$$

From Equations (Eq. V.12) and (Eq. V.13) it follows that

$$\rho=\frac{nc+ns\cdot np}{np\cdot t_r}\cdot\frac{t_{proc\_auth}\cdot t_{proc\_tick}}{t_{proc\_auth}+t_{proc\_tick}} \tag{V.16}$$

Using the obtained equation, we can estimate the probability of denial of service for a user request. This situation is possible when the presence of the number of requests in the queue is so great that the overall request processing time exceeds the time of user residence in the network. Such a queue length corresponds to the ratio of the time of user residence in a network to the average request processing time and is of an order of at least 10^3. For a network that has 1,000 high-mobile subscribers and 9 partners $\rho \approx 0.67$. Substituting this value to Eq. V.16 we obtain the value of probability of adding a 1000th request to the queue

$$P_{1000} = 0.67^{1000}(1-0.67) \approx 10^{-1000} \qquad \text{(V.17)}$$

Thus the probability of the denial of service is very low.

V.3.3 Performance analysis

We have simulated FAP operation to estimate the time of neighbour table creation and the impact of reactive mode of ticket acquisition on the authentication latency. The description of the simulation model used is provided in Annex A. Table V.3 shows parameters used in the simulation. All numbers represent average values for operation execution. Authentication latencies are obtained from experiments described in Section V.2.6.

Table V.3: Parameters used in simulations

Operation	Value
Time for ticket creation	4.48 ms
FAP authentication	30.59 ms
Full authentication	85.33 ms
Propagation delay (inter-domain)	2-24 ms
Propagation delay (intra-domain)	1-2 ms

Introduction of a neighbour table at the FAPS leads to significant reduction in network load. Figure V.17 compares the number of tickets generated and sent to one subscriber using both non-optimized and optimized ticket distribution schemes.

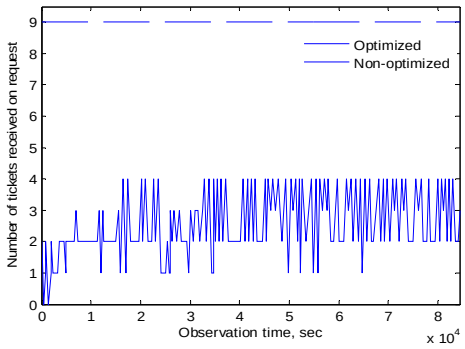


Figure V.17: Number of authentication tickets received by a user in different networks

The duration of the neighbour table creation procedure depends on the number of subscribers and their mobility type. As all networks are in equal conditions, we can average the time of neighbour table creation over all FAPS.

Figure V.18 shows the effect of the number of active subscribers and their mobility type on the duration of reactive mode of FAP operation. As can be seen from this figure, faster clients accelerate the creation of the neighbour table, while with increasing number of users the time of the neighbour table creation increases due to the queue of requests forming at each server.

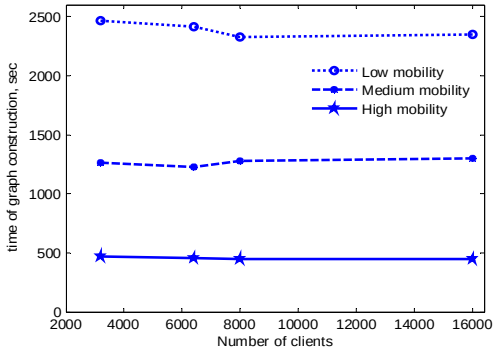


Figure V.18. Time of the neighbour table creation, average for servers

After the neighbour table has been created, the authentication process is executed in proactive mode, when a user has a ticket for a target network before the handover is decided. Figure V.19 presents the evaluation of the average authentication latency for 100 clients, who are subscribers of the same network operator.

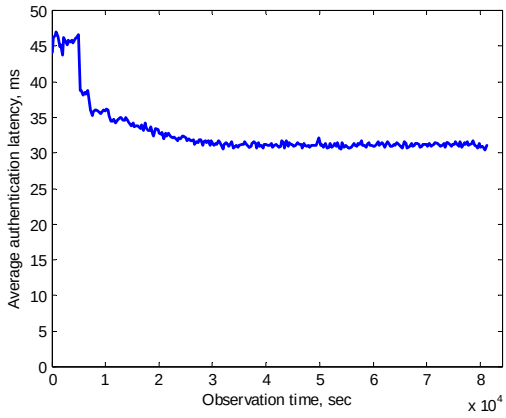


Figure V.19. Average authentication latency for 100 subscribers with low mobility type

The simulation results show that optimization of ticket distribution significantly reduces network load. The reactive mode of FAP operation increases the authentication latency, but it guarantees more efficient operation of the fast authentication protocol in proactive mode.

V.4 FAST RE-AUTHENTICATION PROTOCOL ANALYSIS

A mobile node needs to receive re-authentication tickets after each inter-domain handover. If the user changes networks frequently, the delivery of credentials may cause a significant traffic overhead. In this section we introduce a scheme of authentication ticket distribution that minimizes the network load caused by re-authentication tickets distribution.

V.4.1 Security considerations

The operation of FAP is based on the result of the previous successful strong mutual authentication between the user and a network and does not depend on the used method. The protocol is supposed to be only used for user re-authentications during inter-domain roaming.

The proposed authentication protocol corresponds to requirements formulated in the RFC 4017 [97] to ensure protection of the user, the home and the visited network. Below we provide an analysis of security threats. We assume that due to the nature of wireless network all traffic is visible to a potential attacker.

Ticket interception. During the ticket acquisition phase an attacker may steal a ticket. The interceptor cannot impersonate the valid user with the ticket at the authentication phase because he is unable to decrypt the secret part and does not have enough information to reply to the Challenge message sent by the tFAPS (See Section V.2.5).

Impersonation. The user cannot authenticate a fake network unless the latter has decrypted the ticket. The exchange of Challenge and Response messages in the authentication phase serves for protection against the Man-in-the-Middle attack.

To impersonate the valid user the attacker must have full access to the information kept on the user terminal.

Modification of information. We assume that the user and its home network share some secret and the anchor network signs the Ticket Response message during the ticket acquisition phase. So the user is able to detect data modifications. During the authentication phase the target network can verify the signature of the ticket and, if it is not valid, the tFAPS does not continue authentication.

Discovery of keys. The third party that has revealed the authentication key or a key derived from the key material cannot guess the information used for their generation because all keys are calculated using one-way pseudo-random function. The keys are mutually generated and are not transmitted between the FAPC and the FAPS.

Denial of service attack. At the end of the authentication phase, the malicious node cannot carry out a DoS attack as the Failure message is signed with K_a and the FAPC can authenticate its origin.

Service stealing attack. If the FAPS is compromised or one of the roaming shared keys is exposed then tickets can be created on its behaviour. To privilege its own subscribers and to prevent denial-of-service attacks a network may limit the number of users that can be served in a time period (e.g. per day or per hour) per partner.

V.4.2 Comparison with standard methods

In previous approaches like EAP-TTLS the target network must also communicate with the user's home network to authenticate the user. Table V.4 shows a comparison between the TTLS-MD5 authentication protocol, used for illustration purposes, and the proposed solution.

Table V.4: TTLS and FAP protocol operation comparison

	TTLS-MD5	FAP reactive
Server certificate	Yes	No
RTT MN-target AS	6.5	2
RTT MN-home AS	2	1
RTT target – home AS	1.5	0
Symmetric encryption/decryption	4	6
Asymmetric encryption/decryption	2	0
Signature/verification	1	2

The idea of the method is similar to that used in the Kerberos protocol [123]: to access a service a client presents a ticket issued by the party trusted by both the service holder and the user. The Kerberos was not designed for inter-domain communications and the server is usually an intermediate between two authenticating parties while the proposed protocol facilitates direct communications between the client and the server.

Table V.5: Kerberos and FAP protocol operation comparison

Characteristics	FAP	Kerberos
Information kept by the client	Authentication Ticket	Ticket Granting Ticket, TGS session key
Number of messages exchanged (network access phase)	4	4 (with TGS)+3 (with NAS)
Number of entities involved in network access phase	Client, AP, AS	Client, AP, TGS, NAS
Number of entities involved in secret acquisition phase	Client, Current AS, Home AS	Client, Current AC, TGS, Home AC, TGS
Number of cryptographic operations performed by the server	1 encryption 1 decryption 1 PRF calculation	3 decryption 1 encryption
Number of cryptographic operations performed by the client	1 decryption 1 PRF calculation	2 decryption 1 encryption

The Kerberos protocol requires too much operation to be used for authentication purposes. The extension of Kerberos for inter-domain communication is based on referral tickets, where one network provides a user with the key of a partner. Each server should keep a large quantity of session keys for all its neighbour partners and users. A part of Authenticator is the user's IP address that may have no meaning in the roaming scenario.

If we want to use Kerberos for authentication in the roaming scenario, the service is represented by network access. The mobile node authenticates with AC in the trusted domain and obtains Ticket Granting Ticket (TGT) (3 messages exchanged). After that

it communicates with TGS in order to receive a Client-to-Server Ticket and a Client-to-Server encryption key (4 messages exchanged). Having a ticket, a client starts communication with the server (network access server in our case) exchanging 3 messages.

To access services in a trusted domain, the Kerberos protocol uses referral tickets. The TGS gives a user a TGT and a session key for a TGS in the domain providing required service. The mobile node communicates with this TGS in a foreign domain and receives a session key and a Client-to-Server ticket to access an asked service.

V.4.3 Compared to ticket-based authentication proposals

The proposed Fast re-Authentication Protocol (FAP) implements the concept of recommendation credentials but it differs from approaches described in [124, 125, 126, 127, 128] in some points. Firstly the protocol provides user authentication before any interaction with a visited network, which enforces network protection. As the authentication ticket may be created both by the home and by the visited network the approach tries to extend the mobility region for the mobile user. The proposed authentication ticket does not require any management due to its short validity period. We propose a mobile node-driven authentication scenario, which eliminates communication between different networks.

Table V.6 summarizes the security and management features of the proposed and previous protocols.

Table V.6: Comparison of Ticket-Based Approaches

Characteristics	Related work					FAP
	Hong [124]	Wang [125]	Long [126]	Polito [127]	Ohba [128]	
Communication target – home AS	No	No	No	If cache miss	Yes, reactive	No
Communication target – current AS	Yes	No	No	If cache miss	No	No
PKI	No	Yes	Yes	Private	No	No
Mutual authentication	Yes, optional	Yes	Yes	Yes	Yes	Yes
Related protocol/Level	Mobile IPv6/ Network	Mobile IP/ Network	SSL/ Link, Application	EAP/ Link	EAP/ Link	EAP/ Link
Issuer of credentials	Visited	Home/ Visited	Home	Home	Home /Broker	Home/ Visited
Management of credentials	No	Yes	Yes	Yes	Yes	No
Key material derivation	Yes	No	Yes	Yes	Yes	Yes
RTT	2*/3**	1.5	4.5	2.5	1*/2.5**	2

*Proactive mode, **Reactive mode

The proposed Fast re-Authentication Protocol also implements the concept of recommendation credentials but it differs from approaches described above in some points. First of all, the protocol provides user authentication before any higher layer interaction with a visited network, which enforces network protection. As the authentication ticket may be created by both the home and the current network the approach extends the mobility region for the mobile user. The proposed authentication ticket does not require any management due to its short validity

period. We propose a user terminal-driven authentication scenario, which eliminates communication between different networks. Table V.6 summarizes listed approaches.

V.4.4 Summary

In this section we have presented a Fast re-authentication Protocol for inter-domain roaming. FAP localizes the authentication process, eliminates the need for heavy management of user credentials and minimizes communications between different domains. The aim of the proposed solution is to minimize authentication time and consequently the overall time for inter-domain handover. In-session inter-domain communication is still needed for management and ticket acquisition reasons. However, these interactions are not critical for the handover process. FAP allows mutual generation of key material, which serves to produce session encryption keys.

The protocol is supposed to be implemented for the first authentication in a new target administrative domain. All subsequent authentications within the same domain may be optimized using intra-domain fast re-authentication methods such as described in [14, 109, 111, 114, 117].

The knowledge of the client's neighbourhood of the current network of attachment may be used to reduce the number of tickets generated and sent to each user. If the FAPS knows the current location of its subscriber and it knows which partners adjoin its network, it only generates and sends tickets for these partners.

We have presented the optimized distribution of tickets for fast authentication protocol. The proposed solution reduces network load at the ticket acquisition phase and makes it possible to serve a greater number of highly mobile users. We have introduced the reactive mode of FAP operation, in which a home network creates a neighbour table containing information about the presence of a physical path between its roaming partners.

We have implemented Fast re-Authentication Protocol as a new EAP method to avoid modifications at the access point and minimize modifications on the authenticator side. The aim of our experiments has been to study the performance of the authentication phase of the protocol. In our simulations we estimated the time for neighbour table creation and the impact of reactive mode of ticket acquisition on the authentication latency as functions of the number of subscribers and their type of mobility.

V.5 CHAPTER SUMMARY

In this chapter we have introduced two approaches to improve security related signalling during handover. We have started with a proposal for compound user authentication in a visited network, which addresses the problem of long authentication latency in the scenario where service access authentication is decoupled from network access authentication. The proposed approach makes the authentication to a service transparent for a user. The modus operandi is based on the combination of standard protocols such as 802.1X and, for example, PANA.

The handover process still takes a long time and does not allow real-time applications to run without soft handover support. It has been assumed that handover latency may be reduced by the use of pre-authentication schemes defined for PANA.

Following analysis of the vulnerabilities and performance of inter-domain pre-authentication carried out in [M. Komarova, "Problem Statement for Authentication Signalling Optimization". IEEE 802.21 MIHS Project; DCN 21-07-0387-00-0000, 2007] the pre-authentication approach has been considered costly and non-scalable. Thus a new method for fast authentication has been proposed.

The Fast re-Authentication Protocol eliminates the need for communication between the target and the mobile node's home network during handover execution. The authentication process is based on the use of lightweight authentication tickets containing information about the previous authentication result. Our approach is considered to be independent of the underlying wireless technology and the authentication method implemented in the previous network of attachment.

In order to decrease the number of tickets issued for the mobile node by its home network, the optimized scheme for ticket distribution has been proposed. The neighbour table constructed dynamically by each authority having roaming partners and serving mobile users not only allows pre-authentication signalling optimization but may also serve to provide a mobile user with information for target network selection in a handover.

We implemented the fast re-authentication mechanism on a test platform, which is described in Section V.2.6. Our experiments have shown how the proposed approach can decrease inter-domain authentication latency. As we have implemented the proposed protocol as a new EAP method, it can be easily integrated with the compound link-layer and network-layer authentication approach.

We studied the effectiveness of the proposed ticket distribution scheme by a series of simulations that are described in Annex A.

Chapter VI Trust-based access control architecture

The purpose of this work is to provide a service provider or resource holder the opportunity to evaluate the trustworthiness of each potential client, react to the client's activity by adapting access policies to the actual risk level, and derive user's access rights according to his previous behaviour, recommendations from a third party and the actual circumstances. It is supposed that the system is able to observe and to log the activity of each client and use this information to calculate corresponding trust values. Clients with low trust due to illicit behaviour have limited access to services provided by the network, or they are not even allowed network access. Users are motivated to gain higher trust to enjoy access to a larger set of services with higher quality of service.

Formalization of human understanding of trust may serve to treat user behaviour history better, in order to estimate a risk that serving this user represents to a network, to restrict access for potentially malicious users, and to favour good users. Trust calculation is centralized and is based on the personal observations of the trustor and on recommendations received from other entities. We distinguish the following grades of trust: the entity can be untrusted, which means unknown, the trust value for this entity is not built yet, trusted with many trust levels, and distrusted; it means that the entity has lost the system's trust due to behaviour that is not allowed by security policies.

Access policies in a network that serves different types of clients such as subscribers and guests are determined by

- ✓ The presence of roaming agreements with other authorities;
- ✓ Authorization delegated by roaming partners;
- ✓ The context of interaction and
- ✓ The history of interaction with each particular client.

A trust model that makes use of these components is required for policy enforcement. The proposed trust-based access control mechanism provides a response to the challenges presented by the ubiquitous environment in the following way. We add a dynamic aspect to trust relationship management between service provider entities. If roaming agreements are established between two providers, each of them is able to construct trust to another one, based on the observation of activity of recommended clients. Access policy modification, in order to enforce resource protection, is carried out in an automated and autonomic manner. We provide a trust formalization model with clear dependency between access policies and the obtained trust value. The observation-based trust calculation permits dealing with a long-term history of interaction with each user and restricting or prohibiting access to malicious users. Finally, we provide a mobile user with a simple method for reputation-based service provider selection.

The proposed model operates in three stages: in the first stage, the client authenticates to the service provider; in the second stage, the service provider calculates the trust

value for the authenticated client; and, finally, the obtained trust value is matched against service access policies to determine access rights of the user.

VI.1 MOTIVATION AND REQUIREMENTS

Design of an access control mechanism suitable for use in the ubiquitous mobile environment and allowing automated access policy management becomes a real necessity.

The second purpose of our work is to provide a mobile user with more choice in network selection and to make this selection more reliable. With the development of different wireless technologies and the decreasing cost of wireless network deployment a user's mobility region, equipped with a wireless portable device, grows. The user roams in an environment that is heterogeneous in different senses. Firstly, access networks are based on heterogeneous technologies. Secondly, different administrative domains implement different methods for user authentication and access control. And last but not least, the purpose of networks may differ. The user may be attached to a profit-oriented network; he may be subscribed to a network of some enterprise or an institution, or he can use services provided by a kind of a non-profit network. The mobility of a user is no longer limited by the networks of the home operator and its partner. Figure VI.1 shows our view of the actual mobile environment. We assume that networks belonging to the same authority trust each other.

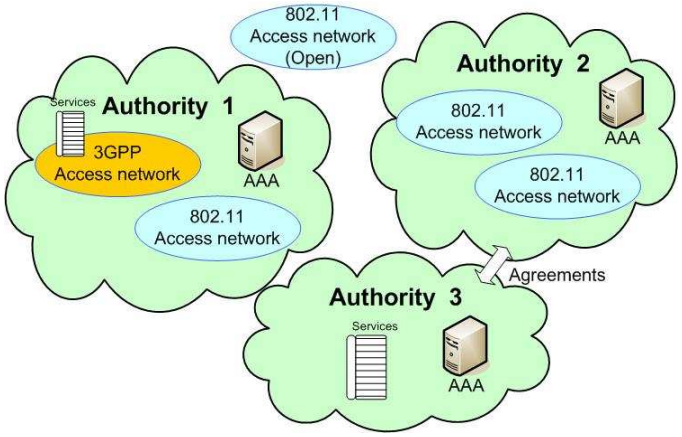


Figure VI.1. Our view of the current situation of trust between service providers and service consumers

Figure VI.2 depicts the scenario in which the proposed model helps to extend the roaming region for a mobile user. According to this scenario, the user was subscribed to two different non-partner federations. Finally, he decided to unsubscribe from one of the federations or the certificate issued by this authority expired and was not renewed. From this moment the user cannot be served by networks belonging to this federation. The implementation of the proposed trust model allows the user who has reached a high trust level in the non-home network to be served by this network even if the user is unsubscribed from the corresponding federation.

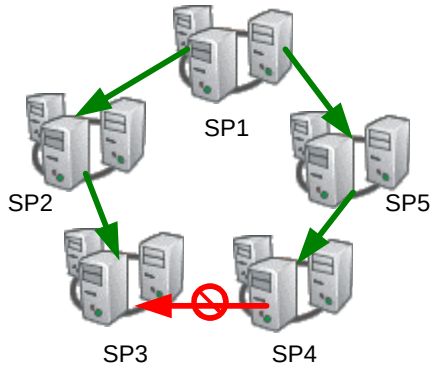


Figure VI.2. The absence of roaming agreements hinders from the ubiquitous mobility

The main idea of the proposed trust-based access control is to provide a user with access to services according to the actual level of trust a network has in him. All operations must be transparent to the user and must not affect authentication performance of this network. This model is designed to manage access to resources in an open environment. The trust-based access control may be implemented for both web resources and authorizing access to networks. In the latter case, the user wishing network access can be the mobile and he cans handover to the network implementing this kind of access control. This case provides strong performance and complex, related requirements.

VI.1.1 User Perspective

Users are motivated to gain higher trust because trusted entities have access to a larger set of services with higher quality of service within one network. The second advantage of the proposed approach is the possibility to gain access to a greater number of networks. Long experience with the network allows a user to roam independently of the presence of recommendations from the third party, the status of roaming agreements, and the degree of trust between the recommending party and the visited network.

Constructing the reputation of service providers on the basis of personal observations affords the user an additional criterion for communication partner selection. In this way, such a procedure may suit a particular user's requirements, and thus the use of unreliable services is avoided.

VI.1.2 Network perspective

From the network's point of view, the proposed trust model also presents several benefits. Existing approaches assume that the visited network serves a client, subscribed to a roaming partner, following successful authentication. But recommended and authorized clients can display bad behaviour, execute illicit actions, infect the internal network entities with viruses etc.

Often trust relations between different authorities are represented by federations of service providers where user identity is understood for each federation member. The assumption made is that a service provider serves all users coming from its roaming partner. If there are many users that show malicious behaviour coming from this part, the trust relations between two networks need to be modified.

Each administrative domain implements its own security policies. The same actions may be permitted in interaction with one service provider and not allowed while interacting with another. The entity does not completely trust recommendations coming from the third party even if the recommender is trusted. In such a way, it is unreliable to make a decision about the trustworthiness of a client based only on recommendations.

VI.2 CONCEPTS AND NOTIONS

VI.2.1 Our understanding of trust

Trust has been defined in different ways in different works, depending on the purposes and usage scenarios for which the trust concept is implemented. To formalize the reason for collaboration between two entities, both trust and reputation models are discussed in the literature.

People will not completely trust somebody based solely on his reputation. We distinguish notions of trust and reputation in the following manner. Trust represents an active and decisive concept: if one entity trusts another entity, the latter is allowed a determined set of actions. Reputation may serve as a source for trust; however, it does not directly define allowed actions. Trust is subjective; reputation, however, is also subjective but is not based on personal observations. Trust always has a clear reason: one entity trusts another on account of some information or experience. For example, in the 802.11i scenario the user (supplicant) trusts the authentication server via a pre-shared secret and he does not trust an access point (authenticator) before performing mutual authentication. After authentication completion the user implicitly trusts the access point via the authentication result and derived session keys. The name of the access point (ESSID and MAC address) is known to the user in advance and ESSID identifies an entity, in particular a network that has some reputation known to the potential user. For the supplicant this reputation represents a reason for association with this access point, but the supplicant does not trust this entity to redirect its traffic prior to completion of mutual authentication, in other words, before trust establishment. Reputation usually comes from an external universe and it does not reflect personal experience of the interested party.

In our understanding, reputation is one reason for establishing trust. We define the mechanism for establishing the relationship between reputation and trust.

Trust is always context-dependent and situation-dependent. In our model the *context* is defined by the user's role (a subscriber, a recommended user or a known user). Each role defines a set of services and resources potentially accessible for the user. The *situation* means the actual relations with partner authorities, the actual set of proposed services and the actual level of risk in the managed environment.

Our trust model is designed to assure more secure interaction between two entities. One of these entities provides various services and another entity requests and consumes services; for that reason, we consider a client-server collaboration model. We define trust in the context of our model as follows.

One entity has a certain degree of trust to another entity and therefore allows it to execute certain actions and to access some resources because the trustee has collaborated with the trustor in the past and the latter was satisfied with the result of this collaboration, or a trusted third party has recommended the trustee. So, the trustor supposes that the trustee will behave satisfactory during the interaction.

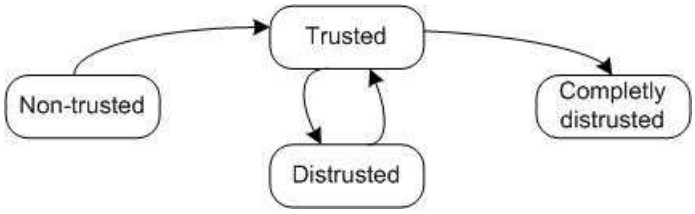


Figure VI.3: Degrees of trust

We distinguish degrees of trust as it is shown in Figure VI.3. An entity may be *non-trusted*, which means that the trust has not yet been formed. Over time and after several interactions an entity becomes *trusted* with a corresponding trust value and, finally, it may become *completely trusted*, *distrusted* or *completely distrusted*. If the entity is completely trusted, it is allowed to perform all actions associated with the given type of entities. The difference between distrusted and completely distrusted states is that a distrusted entity may potentially regain trust while a completely distrusted one will not.

VI.2.2 The agents

In our trust model we consider two types of agents, a service provider (denoted as ***s***) and a service consumer (users, denoted as ***u***). There may be contractual relationships established between service providers, and one service provider may recommend his subscribers to his partner service provider.

During each interaction both agents perform different actions. To simplify the model, we restrict the area of possible actions performed by the actors. An action is denoted by ***a*** and it may result in two values, good (positive) and bad (negative):

$$\textbf{\textit{Action:}}\ a \in \{positive, negative\}.$$

Actions performed by a service provider may include, for example, providing network connectivity, allowing access to data storage or providing various kinds of information. The user considers the action performed as positive if either a provided quality of service corresponds to a declared quality of service or the information provided was correct. Otherwise, the action is considered to have a negative result.

The user may perform a wider and less defined spectrum of actions. The action provided by the user is considered positive if it does not conflict with the service provider's security policies. Otherwise it is considered to be negative.

The agent is characterized by a role and a history of previous interactions. This role varies in different situations. The agent providing services may be a home authority or a visited authority as far as the served user is concerned. The user may be a subscriber, a recommended user or a well-known user as far as the served agent is concerned.

Different agents at the same point in time are not considered to be in the same situation. Agents having the same roles and the same history but interacting at different times find themselves in different situations.

All users may alternate between good and bad behaviour. We define a malicious ("bad") user as a user that does not respect security policies applied by the serving provider or who tries to affect resources or services. A "good" user does not launch illicit actions. We distinguish three scenarios for bad user behaviour:

1. Early bad behaviour (see Figure VI.4, a); in this case the user starts to misbehave from the first visits. It is the trivial case and such a user will quickly become distrusted.
2. Random or occasional bad behaviour (see Figure VI.4, b); users who adapt this pattern of behaviour in general behave as good users and only occasionally display suspect behaviour. For example, sending files infected with viruses may cause such behaviour. Users of this type may regain the trust of the network after several "good" interactions, , but the trust will be earned slower than for a user who was never distrusted.
3. Late or strategic bad behaviour (see Figure VI.4, c). Users belong to this group begin to misbehave after having attained a high trust value and, accordingly, more access rights.

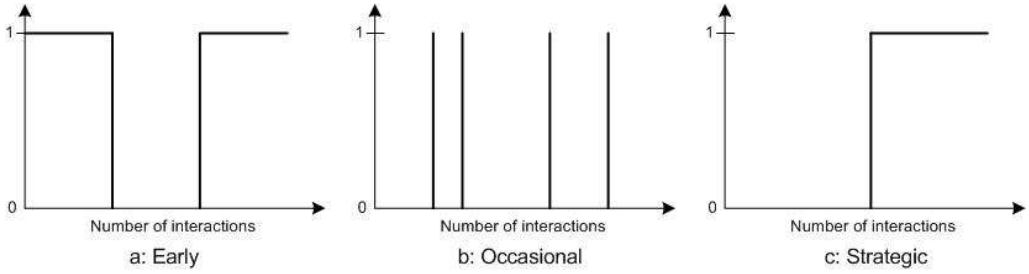


Figure VI.4: Malicious clients with different behavioural patterns

The agent is characterized by a role and history of previous interactions. The role changes in different situations. The agent providing services may be a home authority or a visited authority for a served user. A user may be a subscriber, a recommended or a well-known user for the serving agent.

VI.2.3 Sources of trust

Two main sources of trust are considered in this work:

Personal observations. This is the most trustworthy source of trust. The value of trust is based on the history of a particular client's past behaviour, recorded by a dedicated network entity. If services provided by the access network are located in another network belonging to the same security domain, feedback from these networks is considered a personal observation.

Recommendations are the opinions of trusted authorities concerning a particular agent. In our approach, recommendation expresses the positive opinion and means that the recommended user is considered trustworthy by the recommender. This information is very important if the trustor deals with an unknown user. In this case

it is a single source of trust. Otherwise this source of trust has less influence on decision making about the trustworthiness of the user. The situation in which the recommender is not fair or is not aware of the behaviour of the recommended user is also possible. That is why it is necessary to estimate the trustworthiness of each partner that may recommend users. The estimation of *Reputation of recommenders* is based on observation of recommended clients' behaviour. Information that is difficult or impossible to verify is not used in the proposed trust model. Feedback from partner authorities is considered as a type of information because it is impossible to learn what really happened when the client was served by another network.

VI.3 REQUIREMENTS, ASSUMPTIONS AND LIMITATIONS

We have designed the trust-based access control model to address the problem of trust establishment and trust management between previously unknown agents in an open and dynamic environment. Usage scenarios considered include interactions between peers in an overlay network, interactions between the user and web-services, and interactions between a mobile user and a non-home access network. In the latter scenario, we also consider the handover problem. A mobile user moves among different service providers represented by access networks. The mobile node running a real-time session that is sensitive to the latency of connectivity interruption may initiate a new connection. Consequently, the trust calculation procedure need not increase the authentication time and complicate fast handover execution.

In all considered scenarios different threats are imposed on an agent by other malicious agents. Malicious service providers may offer services that have been advertised or give incorrect of false information. A malicious user may compromise a service provider's security policies, may cause denial of service, obtain access to resources to which they are not authorized, and perform attacks on the service provider's entities and on other users. The proposed mechanism should define a way whereby

- ✓ The service provider can mitigate attacks denying access for potentially dangerous users and protecting fair users.
- ✓ The user is able to distinguish fair and malicious service providers and make a correct choice.

In our model it is assumed that the value of trust to the user is calculated after authentication between the entities that are going to communicate. We also assume that one agent is able to recognize another agent's identity and that the service provider has the means to observe, record and analyze a user's activity.

We assume that each service provider has its own access policies. These policies define sets of services that the user with a particular trust level can access. There should be a clear match between the access policies and the parameters of a trust model. The formalized model should translate access policies into trust in a simple and visible manner.

Just as in the world of human interaction, in the digital world it may take a long time to establish trust, but a relatively short time to lose it. In such a way, it should be

possible to retain long-term history of interactions between agents to avoid the situation in which malicious agents lose trust yet after a short time are able to regain trust and recommence malicious activity. The trust value attributed to the user should depend on the entire past experience. The way to distinguish users making occasional errors and truly malicious users should be defined. To be flexible and adjustable to the changing situation, the trust formalization must take into consideration the observed behaviour of the trustee, the recommendation given by the trusted third party and the reputation of the recommender.

The deployment of any model imposes resource-related limitations. The memory of every system has a limited size; therefore, a long history of interactions between agents should be summarized and retained in an efficient manner. To save on resources (time, battery life and computing power), the trust model must be simple and it must not be computationally heavy.

In development of this trust model the following requirements are taken into consideration:

1. User's reputation should be formalized in a light-weight manner;
2. A service provider should not store a large amount of user-related data;
3. The model should be adaptable for changes in access policies and user behaviour;
4. All parameters of this model should be directly defined by policies set in a natural language and be easy to understand.

The proposed solution must integrate with existing AAA servers and authentication databases, and with the log files of firewalls and intrusion detection systems, if any are present.

VI.4 MODEL FOR SERVICE ACCESS CONTROL

We analyze the following cases of our trust-based access control model deployment:

1. The service provider (it may be an enterprise or a campus network) grants services for free on the basis of membership or a subscription.
2. The service provider has a set of free of charge services and another set of services that it offers for a certain cost.

In both cases the service provider will serve an unknown user only if the latter has a recommendation from the authority that itself has a good reputation. This recommendation may be an X.509 certificate, a user password confirmed by his home network in the authentication exchange, or a service ticket as well. With time the user acquires a reputation in the access network, and lastly, the recommendation from a trusted authority is not necessary for the client with a good experience to be granted access to the free services.

Depending upon the services provided, trust in a user may have many levels, as well as simply two levels (trust/do not trust). Figure VI.5 shows an example of the access control deployment that provides several types of services depending on the level of

trust in a particular user. An unknown user is considered as non-trusted and may be granted a basic non-privileged set of services that may include limited bandwidth and limited possibility of accessing or downloading information. If this user visits the service provider frequently and manifests good behaviour, he becomes firstly a near-trusted and then a trusted client. As the trust level increases the client's access rights also increase. The "bad" or malicious client is considered distrusted and is prohibited from accessing the serving network. Two thresholds define each trust level: the lower and the upper thresholds.

Trust levels	Groups of services	Description
Trusted	S(T3)	Access to specific services
Near-trusted	S(T2)	Internet access, higher speed, higher limit for download
Unknown	S(T1)	Internet access, limited speed, limited download
Distrusted	S(T0)	Access denied

Figure VI.5: Example of service sets and corresponding trust levels

Authorities that are trusted by the service provider are combined into *Contractual groups (CG)*. Each contractual group has a set of agreed services. A user unknown to the serving network is given access to the service set corresponding to the Contractual group to which the recommender belongs. Over time the service set available for a user is either extended or reduced, according to the actual trust value. The reputation of each authority also evolves as a function of the behaviour of recommended clients and of payment for services used. In this model we consider only the behavioural component of this function.

We define the following sets of services provided by the access network: *S(T)* – service set for each trust level *T*, *S(CG)* – service set for a contractual group *CG*.

The user ***u*** has a recommendation from the authority ***m*** and it tries to join the target service provider ***s***. To fix the appropriate service set the access network uses the algorithm shown below:

if $S(T_s(u)) \cap S(CG_m) \neq \emptyset$

then $S' = S(T_s(u)) \cup S(CG_m)$

else $S' = S(T_s(u))$

If the service provider offers a chargeable service, the presence of a Recommendation is mandatory, because the reference is required for a payment source. Definition of payment schemes is outside the scope of this work.

VI.5 TRUST IN A USER: GENERALIZED MODEL

We develop a centralized trust model, in which the entity providing services does not completely trust its partners and relies on its personal observation rather than on

feedback or recommendation from third parties. The trust value for each user is computed before authorizing him to access network resources.

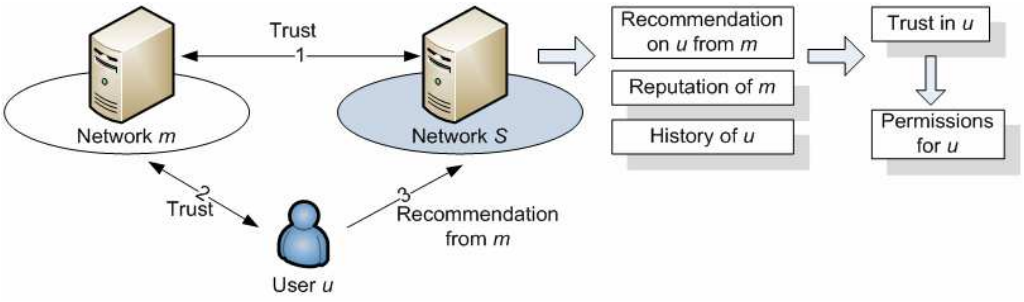


Figure VI.6: General trust construction

In our model, trust is calculated based on the experience the network has with each user, recommendations on the user (e.g. certificate) and the reputation of the entity that has recommended the user (Figure VI.6). Services may be located both in the network managed by the service provider and in its partner network. Based on the evaluated trust value, one of two possible solutions is selected, to allow or block the access to services for a particular user. Feedback from other service providers can be taken into consideration in the following case: an access network may grant access to several services to users authenticated in this network. To construct the history of the user's behaviour, each service sends feedback to the entity responsible for Trust computation.

The proposed trust evaluation approach allows decisions to be made about the user's trustworthiness, taking into account developing trust and a dynamically changing environment. The trust evaluation and definition of the available service set are performed automatically. These procedures are transparent to the user and do not require the intervention of a system administrator.

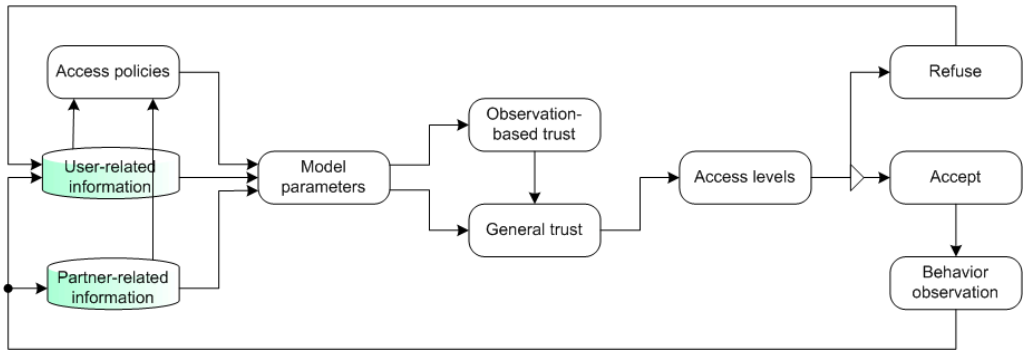


Figure VI.7: Components of the trust model and their combination in the trust calculation procedure

Figure VI.7 depicts the principal components of the proposed trust model, the information flow and interdependency between components. When a user requests access to services, the service provider generates parameters for trust calculation based on the interaction history with this user, information concerning the agent recommending this user, and current access policies. Updated parameters are used to calculate the observation-based trust value that is used to construct a general trust value. The trust value obtained is mapped with access levels defined in advance, and as a result of this mapping the service provider makes a decision to serve or not serve the

user making the request. The activity of an accepted user is observed, analyzed and recorded in the history both of the user and the agent (partner) that has recommended this user.

To determine the trust level and corresponding services to which the user is authorized to access, two evaluation factors are defined, the recommendation from the trusted party and the history-based behaviour observations.

The trust level attributed to the user depends on the actual perceived risk at the moment of user's access request arriving. For example, two users having the same history but asking services from the same network at different moments may be given different access rights.

VI.5.1 Computing general trust

Trust relations are always bilateral and are not symmetrical. The fact that one agent "*s*" trusts another one "*u*" is denoted as $T_s(u)$. Trust values continuously change in the interval [0,1]. Hence, one agent completely trusts another one if $T_s(u)=1$ and completely distrusts a corresponding if $T_s(u)=0$.

We present a formalized model for trust calculation based on the analysis and reasoning provided above. Table VI.1 gives a summary of notions employed; a detailed description is provided in the text.

Table VI.1: Summary of notations employed

Notion	Meaning
s, u	Agent, may be a service provider or a user
$a \in \{positive, negative\}$	Action, during one interaction between agents several actions may be performed
$T_s(u) \in [0,1]$	Trust *s* has in *u*
CG	Contractual group
S(T)	Service set associated with a trust level *T*
S(CG)	Service set associated with a Contractual group *CG*, $S(T) \bigcap S(CG) \neq \emptyset$
$T_s^{(o)}(u) \in [0,1]$	Observation-based trust
$R(m) \in [0,1]$	Reputation of the agent *m*, $m \in \{s\}$
$A(m,u) \in \{0,1\}$	Recommendation (advice) given on the agent *u* by the agent *m*
$\beta \in [0,1]$	Weight of observation-based trust in computation of general trust in a user
tl	Learning time, the period of studying user's behaviour
n_{pos}, n_{neg}	Number of positive and negative experiences with an agent
α(u)	Optimism, determines the rate of observation-based trust earning
K(u)	Tendency, determines the actual maximum achievable value of observation-based trust for an agent *u*

Trust relationships between two agents may be established only if at least one source of trust is available at the moment of collaboration. We consider that the full trust value in the client is formed from values of experience or observation-based trust $T_s^{(o)}(u)$, reputation *R(m)* and recommendation (advice) *A(m,u)* as follows (Eq. VI.18):

$$T_s(u) = \beta \cdot T_s^{(o)}(u) + (1-\beta) \cdot R(m) \cdot A(m,u), \qquad \text{(VI.18)}$$

Experience (*Observation-based trust*) expresses the result of the interaction with the particular user *u* in the past. The service provider itself calculates this value and takes

on real values form zero to one. A detailed explanation and computation model for this component are given in the Section, "Adjustable Observation-based trust model".

Reputation generally shows the common opinion about the trustworthiness of an agent. It may be based on feedback from other agents. In the proposed model, reputation is used to construct trust in an unknown user and it represents the reputation of the agent that has recommended this user. If the agent has a good reputation, the service provider trusts recommendations issued by this recommender, and this trust is based on its reputation. Generally, the serving network does not trust all its partners equally.

Let us take some examples. The username, advertised by a potential client ***u***, may contain a name of his home authority ***m***. This authority may be a known educational institution and the serving network may consider that this user will not be malicious. The network mentioned as the identity provider of the user may be known for its strict security policies and this fact provides a reason for the service provider to grant the user access to resources. For some public open networks such as hotel or airport hotspots, the user location (closeness to an access point) may serve as reputation. For some access points (e.g. the 6^{th} floor of a hotel) the attached terminal belongs to a user who is a client of the hotel with higher probability than for others (the first floor of the hotel). In our model the term "Reputation" means the reputation of an entity recommending the user to a service provider. In the formula for trust calculation, reputation serves as a weight for recommendation.

One agent ***n*** will start to collaborate with another agent ***m*** if the latter has a good reputation. Here collaboration means contract negotiation between ***n*** and ***m***, joining a federation or an agent's ***n*** engagement to serve users that are recommended by the agent ***m***. Generally, this kind of relations will not be established with an agent who has a bad reputation. Since it is difficult to establish a correspondence between "good" or "bad" reputation from an external source and a numerical value, we shall therefore assume that a reputation value is assigned to a new partner in an optimistic way; it has the maximum possible reputation value.

In the formula for trust calculation, reputation serves as a trustworthiness weighting for recommendation. Reputation of the recommender changes over time and it depends on the results of interactions with recommended users. Reputation takes on real values from zero to one. We do not consider negative values because there is no interest to collaborate with an agent with a negative reputation and it is useless to keep and manage exact information about such an agent.

For each of its partners the service provider keeps the number of interactions performed with users recommended by this partner N and the number of interactions considered as successful or positive n_{pos}. These two values compose the *rate of positive recommendation*. After each interaction with a recommended user the rate of positive recommendation is renewed. N' and n'_{pos} denote the updated general number of interactions and the updated number of interactions with positive result correspondently. It may remain the same in a case where there were no negative interactions performed in the past, and it may increase or decrease. In the two last cases the rate of positive recommendation should have effect on the reputation value. The Reputation value for a partner m is recalculated as follows:

$$R'(m) = R(m) + \left(\frac{n'^{(m)}_{pos}}{N^{(m)'}} - \frac{n^{(m)}_{pos}}{N^{(m)}} \right) \quad \text{(VI.19)}$$

The reputation value for a particular partner may be renewed either after each interaction with a user, recommended by him, or after a certain number of interactions. Finally, all partners of the server provider may have different reputations. Figure VI.8 provides an example of one partner's reputation development over time. The starting reputation value is set to "1", and after several negative experiences with recommended users the reputation value decreases, while in the absence of negative experiences the reputation may be restored.

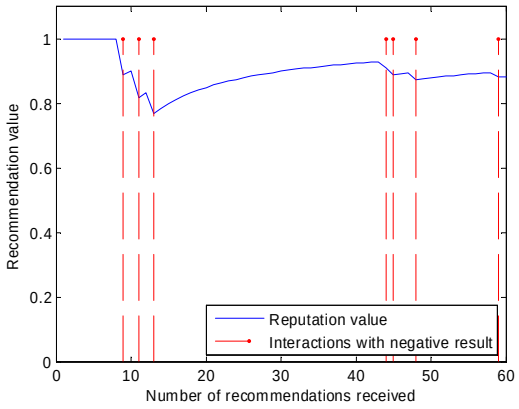


Figure VI.8.Example of recommender's reputation development

In the proposed model reputation already reflects the degree of trust in each recommender. A user may be recommended both by more and by less reputed partners and the service provider takes into consideration the recommendation from the most reputed recommender. We don't aggregate all recommendations associated with the user into a single value in a way similar to that adopted in trust and reputation models for peer-to-peer communications. As recommendations in our model mean that the recommender has a good opinion of the user, the recommendation receiver takes the opinion of the more trusted recommender.

To demonstrate how the resulting recommendation value can be obtained we use the PGP model, in which one certificate may be signed by multiple authorities. The certificate holder is served if the certificate contains a signature of at least one authority trusted for the service provider. The example below (Figure VI.9) shows how the service provider derives the resulting reputation value for a user recommended by more than one service provider's partner. The resulting reputation value in this case corresponds to the maximum value of available recommenders' reputation.

Reputation of recommenders				
0.8	0.3	0.6	0.1	0.9
Resulting reputation				**0.9**

Figure VI.9. Calculating the reputation value for a user, which has more then one recommendation

Recommendation means some direct statement concerning a user presented to a potential service provider by a trusted authority. A digital certificate may be viewed as an example of recommendation information. The way in which a user provides recommendations is outside the scope of this work.

An agent may recommend a user to another agent who is in this situation and is playing the role of a service provider by, for example digital certificate or service ticket issuing, or by confirming a user's identity in the authentication process between the user and the service provider. Recommendation in our model takes on two values – "0", which means the absence of recommendation, and "1" in case of the presence of recommendation. The trust value for an unknown user is computed on the basis of the reputation of the recommender. If the unknown user has no recommendation than he is refused to access resources provided by the trustor.

Previously received recommendations serve to form the value of recommender's reputation as described above. As the recommendation of users that show negative behaviour leads to degradation of the recommender agent's reputation, the agent is motivated to recommend only good users to its partners. To maintain good reputation in partner networks, the agent provides recommendations for a user according to the locally computed observation trust to this user.

To select users worthy of recommendation, a trustor sets up a recommendation threshold RT. If the trust value in the user is greater than this threshold, the authority recommends this user to its partners, otherwise it does not.

These restrictions preserve the recommender from losing its reputation at the partner side. This may happen if the authority recommends unfair clients. Implementation of such a scheme permits an authority to retain a good reputation.

The following example is provided to demonstrate possible implementation of the proposed scheme. Let the agent $\boldsymbol{n}$ be an access network belonging to a federation of Internet Service Providers and the agent $\boldsymbol{m}$ belongs to the same federation and is an Identity Provider of a user $\boldsymbol{u}$. When $\boldsymbol{u}$ is going to use services provided by $\boldsymbol{n}$, the latter verifies his identity with $\boldsymbol{m}$. If u is considered as a "good" user by $\boldsymbol{m}$, it confirms user's identity. Otherwise it indicates to $\boldsymbol{n}$ that user's authorization has failed. If recommendation is given in form of a digital certificate, the certificate issuer may revoke this certificate for external use and while it remains valid for internal use.

VI.5.2 Trust development

In our model, the influence of different sources of trust on the final trust value develops over time. When dealing with an unknown user, the service provider has insufficient information to estimate the trustworthiness of this user. That is why the trust calculation relies mostly on recommendations received form trusted partners and depends on their reputation calculated by the service holder authority. In time a number of interactions may take place between the previously unknown user and the network providing it with information about user's behaviour. When the trust value

for a user with a certain history of interaction is calculated, the influence of the personal observation on the final trust value increases and, finally, trust is calculated based on personal observations rather than on reputation of recommenders.

In Eq. VI.18 the influence of each trust source on the final trust value is expressed by weight β (observation-based trust) and (1- β) for reputation of recommenders. This parameter is defined as a function of the history of interaction with each user in the following way: for an unknown user the recommendation is more important than experience in the formula for trust calculation.

At the beginning of interaction between two agents unknown to each other, the service provider collaborates only with users recommended by its partners. Then, during the period called the *learning time* ***tl*** the allowed service set is determined both by the user's reputation and by the presence of a recommendation for him. Finally, the trust to a well-known user depends only on his pas behaviour. The notion of *learning time* includes not only the time passed from the very first visit of the user to the present moment but also the number of interactions performed between the user and the service provider. It is necessary to distinguish between a user who performed one interaction during a month and one that performed thirty interactions during the same period. The value of the weight for trust calculation is obtained for each session, using the following equation (Eq.VI.20):

$$\beta = \begin{cases} \frac{n_{vis}}{tl}, \text{if } \frac{n_{vis}}{tl} \leq 1 \\ 1, \quad \text{else} \end{cases}; \qquad \text{(VI.20)}$$

where n_{vis} is the number of interactions (sessions) performed between the service provider and the user. At the beginning of collaboration with an unknown user the service provider has no idea about the trustworthiness of this user and therefore it needs to trust recommendations as the single source available of trust in the user. Interactions performed during the learning time period serve to construct the behaviour pattern of the user. Negative experience gathered may be either occasional events or may characterize intentional attack attempts. The experience collected is used in the third phase of interaction with the user when observation-based trust determines general trust in the user. If the service provider offers chargeable services the recommendation is still necessary to confirm the ability of the user to pay for consumed services.

It can be seen from Figure VI.10 that the more the user interacts with the service provider the faster it becomes independent in terms of obtaining access to its services from its identity provider or another agent providing recommendations. In this figure, three graphs represent development of β parameter for three users. The first of them solicits services every day, the second does so less frequently and the last communicates with the concerned service provider only occasionally. The learning time is equal for all users and it is set to 60 days.

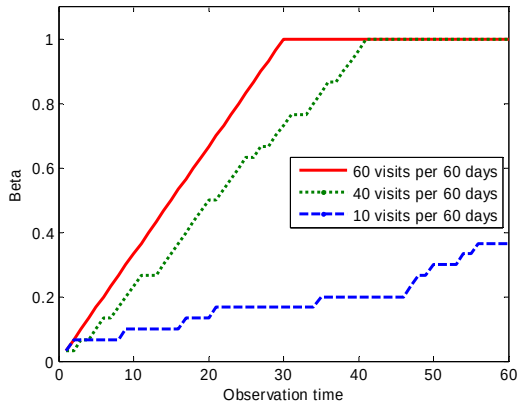
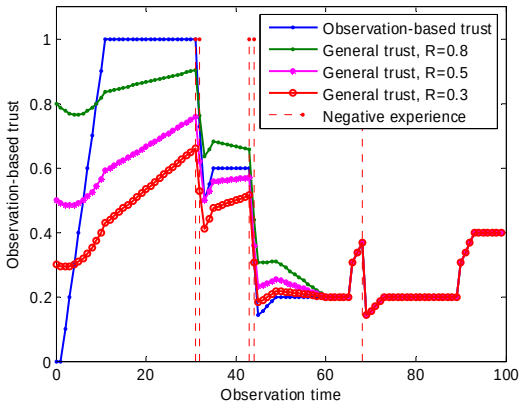


Figure VI.10. Example of experience weight (β) evolution for users with different frequency of visits.

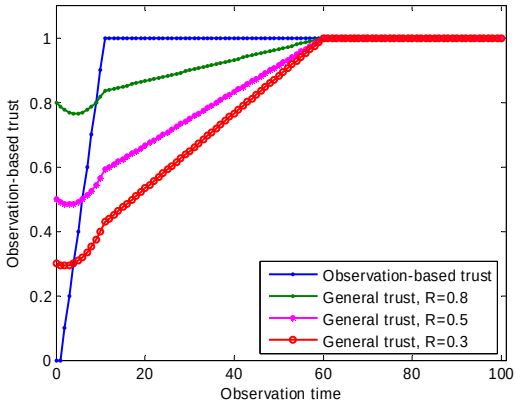
When the service provider collaborates with a user recommended by a new partner, this partner has a good reputation, and hence the user is supposed to be trustworthy starting with the first visit. After a time, the user either remains trustworthy if his behaviour is considered positive or he loses a certain degree of trust after each interaction that has a negative result.

An unknown user also may be recommended by an agent who is well known to the target service provider. In this case the reputation of the recommender depends on the past history of interactions with suggested users. The degree of trust of this user is set in accordance with the reputation value of the recommender.

Figure VI.11 gives two examples of trust earning by a user. Each example depicts both the observation-based and the general trust. In the first case (Figure VI.11, a) the user displays only positive behaviour. The rate of gaining trust is determined only by the reputation of the recommender. In the second case (Figure VI.11, b) several interactions with the user are considered negative. During the learning period the resulting value of general trust may be higher than the value of observation-based trust if the user's recommender has a good reputation.



a



b

Figure VI.11. Effect of observation-based trust and recommender's reputation on forming a general trust value

VI.6 ADJUSTABLE OBSERVATION-BASED TRUST MODEL

Observation-based trust is the most reliable source of information available for a service provider. The proposed model is designed for automated decision-making concerning the trustworthiness of each particular user, based on recorded past experience and service access policies. The model allows correction of access policies according to the actual level of risk perceived by the service provider's system in order to minimize the rate of the observed negative behaviour of users. Trust evolves over time and users considered non-trustworthy in the past may be forgiven, based upon the access policies. This kind of trust may be calculated for subscribers of the service provider as well as for its guests, if representation of their credentials permits unique user identification.

To simplify explanation, observation-based trust is called "trust" in this chapter, since only this source of trust is considered here.

VI.6.1 Model description

The service provider is motivated to grant access to users. That is why expectation for any unknown user is optimistic. The unknown user is presumed not to be a malicious one and he is granted the minimal trust value sufficient to access the network and is able to attain the maximum possible trust value and access the maximum service set. The minimum trust value corresponds to the basic non-privileged set of services that may be characterized by a limited bandwidth and a limited possibility of access or download of information. If this user frequents the network and demonstrates good behaviour, he becomes a trusted one.

Trust takes on only positive values varying between 0 and 1. Negative values of trust are useful in distributed trust and reputation models, in which trust calculation is based on feedback on one agent's activity from other agents. In the proposed model the user is considered to be distrusted and the service provider does not serve him if the corresponding trust value is equal to or less than zero. In this situation the exact negative trust value has no importance and does not influence decision making about the trustworthiness of this user.

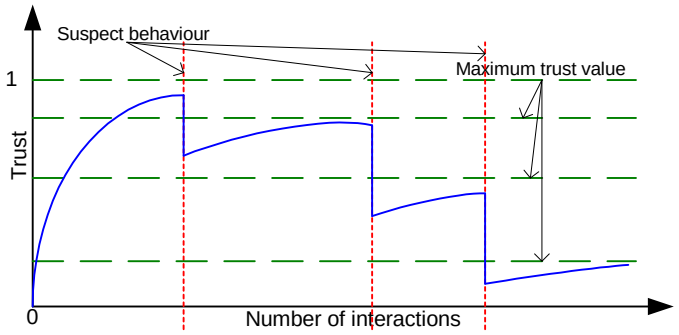


Figure VI.12. Desired development of the trust value to the user over the time

Figure VI.12 shows the desired development of service provider's trust to a user over time. The trust value becomes monotonous for a "good" user until it reaches its maximum possible value. The rate of trust earning depends on access policies of the service provider and the risk level in the environment. The model must react to the suspect behaviour of the user while decreasing the maximum reachable trust value and the speed of the trust growth.

The trust calculation is based on parameters derived from network/service access policies, the number of trust levels and user-related information. *Network/service access policies* used for trust computing include:

1. User u becomes completely trusted ($T_s^{(o)}(u)=1$) after continuous n_{beg_trust} visits with observed good behaviour.
2. User u becomes distrusted ($T_s^{(o)}(u)=0$) after n_{stop_trust} visits with observed bad behaviour.
3. The distrusted user is forgiven with the loss of one trust level after $t_{forgive}$ conditional days.
4. There are m trust levels, each level has an upper threshold Tu:

Level 0: distrusted: $T_s^{(o)}(u) = 0$;

Level 1: $T_s^{(o)}(u) \in (0, Tu_1]$;

...

Level i: $T_s^{(o)}(u) \in (Tu_{i-1}, Tu_i]$

...

Level m-1: $T_s^{(o)}(u) \in (Tu_{m-2}, 1]$

5. The service provider makes access policies stricter if the rate of negative behaviour over all users is greater than or equal to $N_{\max}$. Making access policies stricter means that the value of the parameter n_{beg_trust} is increased and the value of the parameter n_{stop_trust} is decreased.

To obtain an accurate trust evaluation of the user, the service provider retains the following *user-related information:*

1. The number of positive experiences with the user $n_{pos}(u)$; if a user respects network access policy, his behaviour during this visit is considered as good and the number of positive experiences simply incremented.
2. The number of negative experiences in collaboration with the user $n_{neg}(u)$;
3. The number of times the user was distrusted $n_{distr}(u)$;
4. The time label indicating the distrust lifetime and corresponding to the moment when the user may be forgiven $t_{forgive}(u)$.
5. Boolean variable $f(u)$ that indicates whether the user can be forgiven ($f(u)=1$) or whether the client is completely distrusted ($f(u)=0$).
6. The time of the first interaction with the user $t_{interact}$ is used to determine whether the learning period is finished.

To update user-related information, a very simple procedure is used. If the user has demonstrated good behaviour during a visit, this visit is considered a positive experience and the number of positive experiences n_{pos} is incremented. Otherwise this visit is considered a negative experience and the number of negative experiences n_{neg} is incremented. In certain implementation scenarios an interaction with a user is viewed as an atomic transaction while in others several sessions with different results may be hold at the same time. In this case different penalty weights ω_i should be defined for each i^{th} type of misbehaviour (access policies violation).

$$n_{neg}(u) = n_{neg}(u) + \sum_{i=1}^{K} \omega_i , \qquad \text{(VI.21)}$$

where K is the number of access policies violations. The sum $\sum_{i=1}^{K} \omega_i$ may be greater than one.

VI.6.2 Trust formula

Upon an access request from the user, the service provider calculates the updated value of trust according to (Eq.VI.18). $T_s^{(o)}(u)$ formalize the development of trust in the user the linear model has been chosen. All parameters of the proposed linear model are defined by access policies and past experience with the particular user. The use of a linear model for access control has several advantages as compared with non-linear models such as those described in the literature [75, 76]:

1. The main advantage is its simplicity and clarity of understanding; any change in the linear model behaviour or parameters is easy to interpret;
2. Operations performed are not computationally heavy and they do not significantly increase overall authentication and authorization delay;
3. Formalization of the described model with a non-linear model gives the same results;
4. To obtain the same accuracy for the estimate of user's trustworthiness, a non-linear model requires more input parameters [75, 76, 101, 102];
5. It is a known empirical fact that simple liner models often have an advantage in predictive power over more complex non-linear models.

For a "good" user the value of trust grows linearly with an increasing number of visits and reaches the maximum value equal to one. In order to calculate the trust value for the user a discrete formula is defined as follows:

$$T_s^{(o)}(u) = \begin{cases} \alpha(u) \cdot n_{pos}(u), & T_s^{(o)} \le k(u) \\ k(u), & otherwise \end{cases} \qquad \text{(VI.22)}$$

Where $T_s^{(o)}$ denotes the observation-based trust for a user in a particular moment t, $\alpha(u)$ is a parameter of model called "optimism" and another parameter $k(u)$, "tendency", expresses the maximum trust possible to earn for the user with the given history and with respect to actual access policies.

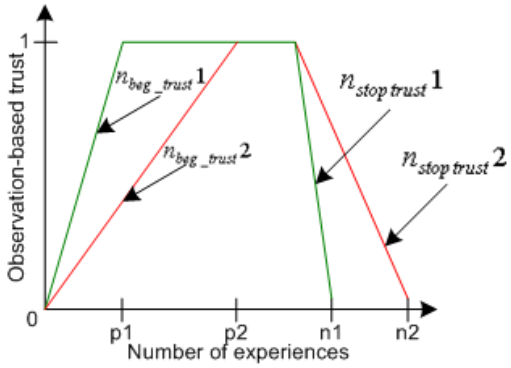


Figure VI.13: Effect of different policies values on trust evolution

It can be seen from Figure VI.13 that trust in the same user may change if access policy changes. With policies $n_{beg\_trust}\,1$ and $n_{stop\_trust}\,1$ the user gains and loses trust more rapidly than in the case of application of policies $n_{beg\_trust}\,2$ and $n_{stop\_trust}\,2$. Values *p1* and *p2* correspond to the number of positive experiences gathered from the

beginning of cooperation between the user and the service provider while *n1* and *n2* correspond to the general number of recorded negative experiences in the overall number of recorded experiences. Hence, two users having the same interaction history (for example, maximum trust gained and after that *n1* negative interactions were performed) but soliciting services where different policies are in force will obtain a different degree of trust. If n_{stop_trust} *1* is applied the user will be distrusted, otherwise the same user may continue to consume services.

VI.6.3 Optimism and tendency

The number of positive and negative experiences defines the trust value for the user, the optimism parameter α and the tendency parameter *k*. The former expresses the rate of trust earning and the latter corresponds to the maximum value that user trust can actually reach.

The *Optimism* parameter expresses the speed of earning trust by the user and it is represented by the tangent of the angle between the line corresponding to trust evolution and the time axis (see Figure VI.14). The upper threshold of the trust level *Tu* is the maximum trust that can be reached by the user. For a "good" user, who has never been distrusted *Tu=1*, optimism is defined by the number of positive interactions n_{beg_trust} that have to be initiated and performed by the user so as to reach the maximum trust, and the number of negative experiences gathered by the service provider during collaboration with this user:

$$\alpha(u) = \frac{Tu(n_{distr}(u))}{n_{beg\,trust} + n_{neg}(u)} \qquad \textbf{(VI.23)}$$

Figure VI.14 illustrates an example of the optimism parameter calculation for a "good" user that did not break any access policy or rule during all past visits. The less the n_{beg_trust} parameter the more the value of optimism ($\alpha_1(u)$ corresponds to $n_{beg\_trust_1}$) and the less time the user will spend to gain the maximum possible trust.

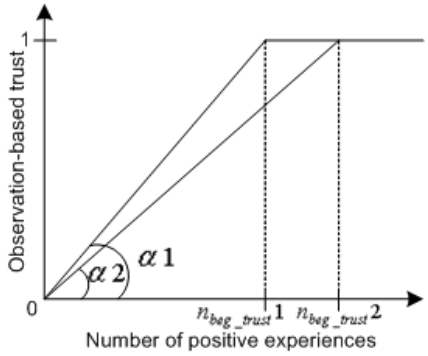
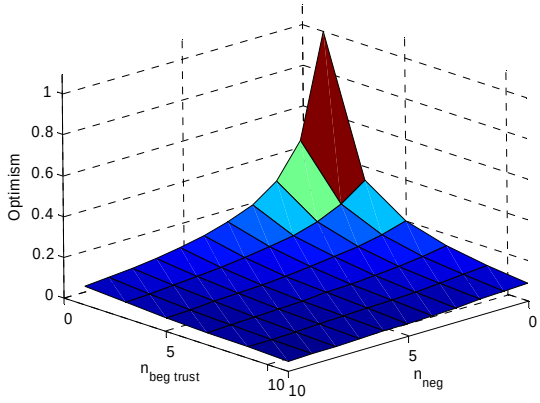
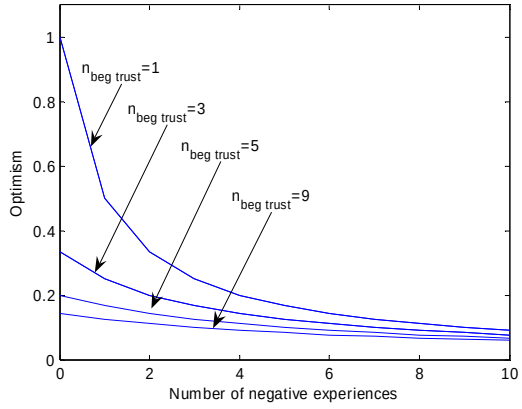


Figure VI.14.Optimism parameter for a "good" client

As was mentioned above, with respect to the actual upper trust threshold set for the user, optimism is defined by two factors: the policy determining the number of positive experiences to gain trust and the actual number of interactions between the user and the server provider considered to be negative. Optimism variation is shown in Figure VI.15.



a



b

Figure VI.15. (a) Optimism as a function of the number of negative experiences, (b) 3D presentation of optimism evaluation against access policy and user-related history

It can be seen that the value of the optimism parameter changes more rapidly after the early negative interactions with the user and it remains at the same level for the user that has corresponding number of negative experiences close to n_{stop_trust} .

The *Tendency* parameter k is introduced in order to regulate the maximum trust value, actually achievable by the user:

$$k(u) = Tu(n_{distr}(u)) \cdot \left(1 - \frac{n_{neg}(u)}{n_{stop\ trust}}\right) \qquad \text{(VI.24)}$$

The difference between tendency *k(u)* and upper trust threshold $Tu(n_{distr}(u))$ lies in the following. $Tu(n_{distr}(u))$ is the maximum achievable trust value and is based on the whole history of collaboration with this user. Tendency *k(u)* also represents the limitation of a trust value, but it is based on local history. For example, the maximum achievable trust value for the user is one, but after several visits with negative behaviour he may reach only trust level according to *k(u)*. After a certain number of

visits with good behaviour the user may regain the possibility to reach trust value greater than $k(u)$, but it is still limited by the value of $Tu(n_{distr}(u))$.

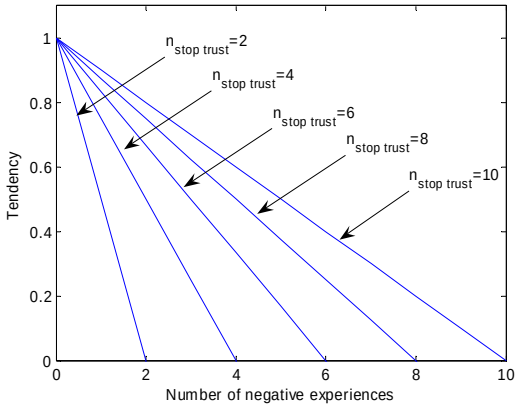


Figure VI.16. Tendency as a function of the number of negative experiences

The value of tendency changes linearly (see Figure VI.16) according to the number of negative experiences associated with the user and the speed of degradation of this function depends on the value set for the policy n_{stop_trust}. The procedure for reestablishment of the tendency value is described in the section "Memory model".

An example of trust variation for a user who shows different behaviour during interaction with the service provider is given in Figure VI.17. A user loses trust depending on the number of negative experiences, and after several positive experiences he regains trust, however, the new trust value is limited by the tendency parameter. Each negative experience decreases the maximum achievable trust value. If the user shows only acceptable behaviour during a series of consequent visits over a defined time period, he can regain the initially set maximum trust value. This procedure is described further in this chapter.

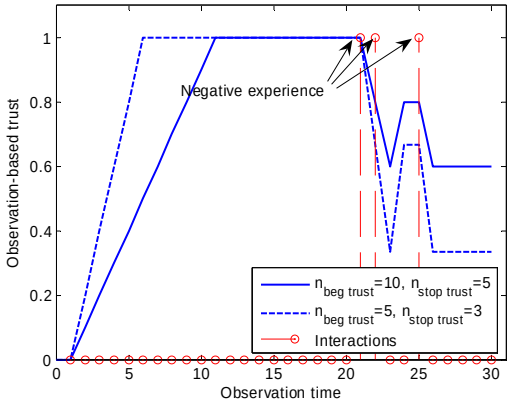


Figure VI.17: Effect of negative experiences and policies on the trust value

The number of positive experiences is not included in the calculation of parameters to prevent attacks on the part of strategic users (as described in Section VI.3 "Assumptions and requirements"). For the strategic user the number of positive

experiences may compensate for the number of negative experiences. In this case attacks held by this user will remain unpunished.

VI.6.4 The memory model and forgiving (past interactions history)

The appropriate model for retaining the interaction history between two agents should be designed taking into account the limitation of memory size dedicated to store history-related data and the timing factor. The timing factor is very important because more recent events must have more influence on the decision about the trustworthiness of the user. However, information about past behaviour should also be taken into account. Generally the history is represented by a sequence of single events [129], [130]. The number of events to retain remains the question. A very long history allows more accurate trust estimation but it requires more processing time and more storage space. A short history lets past bad experiences be forgotten, with malicious users thereby quickly regaining a high trust level. Another problem related to this type of history organization is the aggregation of events in order to compute the actual trust value. In order to represent the varying relevance of events that occurred at different times for the current trustworthiness of an agent, various solutions are proposed in the literature.

Instead of using memory windows, fading memory or forgetting factors proposed in [75,103,85], we keep the history of interactions in scalar variables. The number of positive and negative experiences changes over time due to users' dynamic behaviour and to the system forgetting old experience. However, old experience does not mean obsolete and useless experience. The proposed memory model allows retaining information for long-term observation history and carrying out a more accurate trust evaluation.

In our model we implement different forgetting models for positive and negative experiences. It is necessary to distinguish between the user who was distrusted in the past and the user that was never distrusted. Trust models proposed earlier do not permit this kind of distinction. Proposed forgetting mechanisms are the same for both positive and negative history.

With the proposed trust model the user becomes distrusted after several visits when he has manifested negative behaviour. We define a mechanism for forgiving distrusted users in our trust model. The distrusted user may be forgiven after a certain period of time $t_{forgive}$, defined by the administrator of the system. In general cases, the forgiving period may be defined either on a per-user or per-role basis. It is defined by the actual values of access control policies, risk level and the number of times the user has been deemed distrusted. The service provider may either keep this information in the form of the table of correspondence or compute a function $f(policy, risk, n_{distrust})$ each time the forgiving time has to be defined.

Let us give an example. The service provider has defined four levels of trust with corresponding threshold trust values for each level *{unknown (0, 0.37], near trusted (0.37, 0.63], trusted (0.63, 0.8], completely trusted (0.8, 1]}*. For a user that has never been considered distrusted the maximum achievable trust value is *1*; for one that has been penalized once the maximum achievable trust value is *0.8*; if it has been penalized twice, the maximum achievable trust value is *0.63,* and after the third penalty the user cannot be forgiven. After having been forgiven, the user loses one trust level. For instance, if the maximum potentially achievable trust value before the trust lost was

"*trusted*", then the maximum achievable trust value will be set to "*near trusted*" for the forgiven user.

The system "forgets" the number of positive and negative experiences of dealing with the forgiven user but keeps the number of times this user was distrusted. If the forgiven user behaves well the trust value grows with less optimism for him than for a user that was never considered a distrusted agent. The maximum achievable value of trust Tu depends on the defined number of trust levels m and the number of fatal errors $n_{distrust}$, when access to the network was forbidden him. The algorithm of user forgiving is defined as follows

```
If access request received
  If t(now)≥t_forgiv
      Tu= Tu(m-n_distrust);
      n_pos=0, n_neg=0;
      Serve this client;
  Else deny access
```

Figure VI.18 shows an example of trust development for a strategic bad user. This user starts to perform malicious actions after gaining the maximum possible trust value, aiming to damage more important resources. Each time the user gets distrusted and forgiven the maximum achievable trust decreases. The value of the forgiving lifetime depends on implementation and may be defined by statistical observation. For illustration purposes we have chosen the forgiven time of twenty days. The blue solid line represents trust value to the user and the dashed red lines represent trust levels' thresholds defined by the service provider.

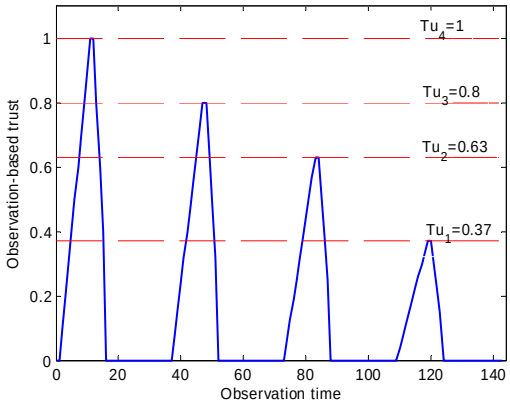


Figure VI.18: Development of trust for a strategic bad user

For a user that has certain degree of trust and, thus is allowed to access services, the forgiving time changes meaning. For example, the agent having the maximum trust value corresponding to "*near-trusted*" has the possibility to become "*completely trusted*" due to his good behaviour shown in a series of consequent interactions. The number of recorded negative experiences may be decreased if the number of consequent positive experiences is greater than the n_{visit} parameter. The service provider updates the user-related information after each interaction completed as described below.

If *behaviour=="positive"*

$n_{pos}=n_{pos}+1;$

If $(t(now)>t_{forgiv})\&(n_{pos}>n_{visit}+n_{beg\_trust})$

If $n_{neg}>0,$ **then** $n_{neg}=n_{neg}-1;$

If *behaviour=="negative"*

If $n_{neg}<n_{stop\_trust}$

then $n_{neg}= n_{neg}+\omega_i;$

If $n_{pos}>n_{beg\_trust},$ $n_{pos}=n_{beg\_trust}.$

If $n_{neg}=n_{stop\_trust}$

then $n_{distrust}= n_{distrust}+1;$

If $n_{distrust}= m,$ *f=false* (can not be forgiven);

$t_{forgive}=t(now)+t_{forgive}.$

Figure VI.19 shows how a user having some bad experience in earlier history is able to regain the trust of the service provider. To illustrate the process of trust earning the number of interactions with a positive result is set to 10. The user that was not distrusted in the past (

Figure VI.19) may over time become completely trusted by the collaborating agent. In case of dealing with a user who was distrusted in the past, the time and the number of interactions needed to regain one trust level are greater than in the previous case. The reason for analyzing not only the number of visits but also the time interval during which these interactions were performed is to make sure that the user has changed the behavioural pattern. The strategic attacker may make a significant number of visits with positive results in a short period of time in order to regain trust quickly. We consider that a user motivated to use the proposed services in the future will show homogeneous behaviour over a long time interval.

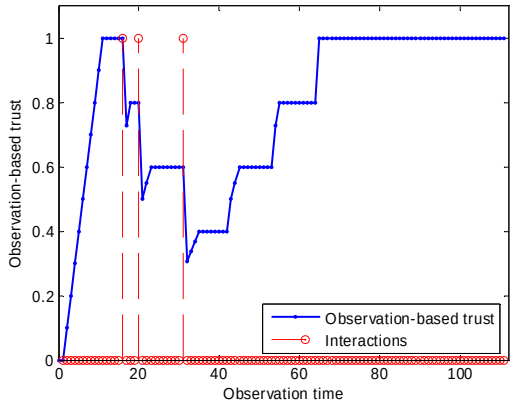


Figure VI.19 Regaining trust by a user

We have evaluated the effectiveness of the proposed memory model in terms of interaction history storage and the performance of the computational model. The proposed memory organization enables the service provider to keep long-term history for each user in only five variables, and operations performed for updating history are linear. Access rights attributed to each user change automatically with

policy changes. As distinct from memory models proposed earlier in the literature, where long-term user behaviour observation is either aggregated into a single value or is kept for limited time period, our presentation of a user's history enables a system to contain comprehensive information about past experience with each user.

VI.6.5 Adapting access policies

In the proposed trust model we use the concept of risk to adapt server provider's access policies to changing environment. At each moment, the ***risk* value** is defined as the ratio of the number of recorded negative experiences calculated for all N users that are allowed to access services, to the overall number of sessions performed with these users:

$$risk = \frac{\sum_{i=1}^{N} n_{neg}(i)}{\sum_{i=1}^{N} (n_{neg}(i) + n_{pos}(i))} \qquad \text{(VI.25)}$$

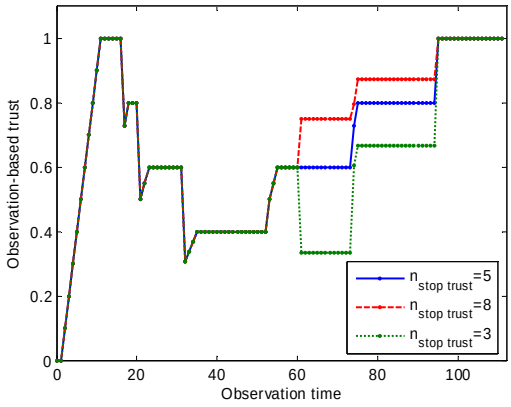
The parameters involved in trust calculation depend on access policies that may change according to the actual risk level. Increasing the number of positive experiences needed to reach the maximum trust value enforces protection against early bad users, and thus they can only cause limited damage to the resources set. Nevertheless, under these circumstances strategic bad users are still able to gain maximum trust from the service provider and consequently privileged access to critical resources. To decrease the negative impact that these users' actions can have on the service provider, the policy corresponding to the number of negative interactions performed with a user needed to lose trust should be decreased. To manage access policies the service provider defines several negative rate thresholds thr_rate_i and corresponding values Δbeg_i and $\Delta stop_i$ by which the policies will change in case the actual negative rate exceeds the given negative rate threshold.

Access policies (n_{beg_trust} and n_{stop_trust}) change according to the following rules:

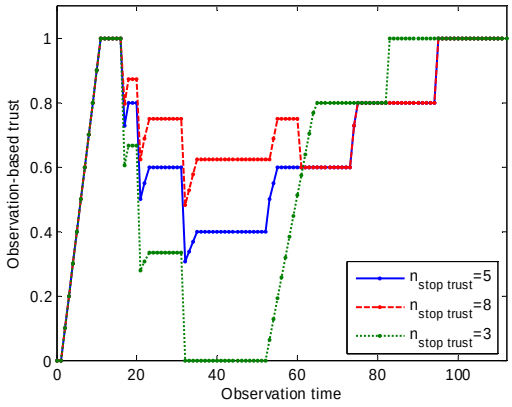
If $rate_{neg} < thr\_rate_1$
then $n_{beg\_trust} = initial\ value$
$n_{stop\_trust} = initial\ value;$
if $thr\_rate_2 > rate_{neg} > thr\_rate_1$
then $n_{beg\_trust} = n_{beg\_trust} + \Delta beg_1$
$n_{stop\_trust} = n_{stop\_trust} - \Delta stop_1;$
…
If $rate_{neg} > thr\_rate_p$
then $n_{beg\_trust} = n_{beg\_trust} + \Delta beg_p$
$n_{stop\_trust} = 1.$

We provide some examples (see Figure VI.20) to illustrate how the access policy variation influences the trust awarded to a user and, accordingly, his access rights. Each user's behaviour has an effect on the risk level in the operational environment. Even an agent who shows only positive behaviour may lose the trust of the service provider due to the high risk level observed.

These two figures depict trust earning by a user who manifests the same behaviour. In the given example only the policy defining the number of interactions with a negative result to lose trust n_{stop_trust} changes. Figure VI.20 presents forming trust in a previously unknown user under different conditions. In the first case (Figure VI.20, a) we investigate the effect of n_{stop_trust} variance in the first part of the history (from the first to the sixtieth visit) on the resulting trust value. It can be seen that when the stricter policy ($n_{stop\_trust}=3$) is applied, the user is not permitted to access services, while with the different policy value ($n_{stop\_trust}=8$) he attains quite a high trust level. In the second case (Figure VI.20, b), the influence of the access policy on the trust evaluation was studied for the already known user (sixtieth and later visits). When the policy changes, the user either loses or acquires a higher trust level. The *tendency* parameter and the maximum reachable trust *Tu* changes correspondingly.



a



b

Figure VI.20: Different access policies and their effect on the trust earning process

If the model with static access policies is used, the system is not able to react to attacks in an automated manner. If the negative rate begins to increase, the system applying dynamic access policies is able to stop strategic users at the beginning of the attack.

VI.7 ANALYSIS AND COMPARISON

Trust establishment plays a key role in today's ubiquitous communications. That is why the construction of trust-based access control policies has been addressed in academic works in the last few years.

Implementation of a trust model that is designed in mobile environment has resource-related limitations such as memory size and computational complexity. We use these limitations as criteria to compare trust models proposed in literature and our model. Other important properties of a trust model include self-adapting and fast reaction to behavioural patterns change. In the comparison provided below we consider neither trust models designed specifically for interactions in peer-to-peer or grid networks nor trust models for domain policies mapping due to their computational complexity that do not address requirements for fast decision making. We concentrate on the recently proposed trust-based solutions for access control and compare their characteristics in Table VI.2. The principle goal of this comparison was to analyse what trust models may be used in an access control framework. In Table VI.2 blue cells correspond to the model's characteristics that make these models suitable to be integrated in the studied scenario, characteristics that make a model unsuitable are marked in orange. Further study is required to determine the impact of characteristics marked in white.

Our trust model uses a memory model that differs from memory and forgetting models proposed in the literature. It allows a network to retain long-term user-related history using only four scalar variables. The originality consists in the possibility of access policies adaptation to the changing conditions of functioning.

The main implementation difficulty related to trust models proposed earlier consists in the necessity for model parameters selection. These parameters are not directly defined by access policies. Figure VI.21 illustrates the trust evaluation in the Beta Reputation System [75], Giang's trust model [129] and our model on a simple example. These models were chosen for comparison because they are designed to be implemented in the same scenario as we consider. We evaluated trust to a user that performed 100 interactions with the studied network. Interactions from 49 to 52 were negative and all the rest were positive. Parameters for the referenced models were set in order to allow a user to gain the trust value of "1" and to be penalized for negative experiences.

Table VI.2: Comparison of Trust-based Access Control models

Characteristics	Pho Duc Giang 2007 [129]	Charkraboty 2006 [130]	Dimmock 2004 [81]	Tchepnda 2006 [101]	Our model
Trust sources	Experience Recommenda-tion	Experience, knowledge recommendation	Credentials, observation, recommendations	History, recommendations, ability to retaliate	Experience, Reputation, recommendation
Trust value	Scalar, [0,1]	Vector of range [-1,1]	Predicate (belief, disbelief)	Scalar, [-1,1]	Scalar, [0,1]
Complexity	Exponential	Exponential	Implementation-dependent	Exponential	Linear
Trust object	Users	Users	Users	Users, partners	Users, partners
Experience granularity	Successful unsuccessfull	Trust positive, trust negative, trust neutral	Context-dependent	Success, failure	Positive, negative
History retaining	Sliding window	Weighted window	Implementation-dependent	Recurrent calculation, no history	Set of scalars
Number of variable to retain history	Window length,	Length of time interval, non-negative weights for each element	Implementation-dependent	-	4 variables
Recommendation	Weighted average	Weighted average	Probability expectation value	Weighted average	Maximum value
Fixed/evaluated parameters	Fixed, static	Fixed, static	Turning, static	Threshold-based, static	evaluated, dynamic
Policies	Trust levels	Permissions and restrictions	Trust values to cost comparing	Context dependent	Trust levels, trust model parameters
Self-adapting policies	No	No	No	No	Yes

In the Giang's model the malicious user regains a high level of trust just after stopping to behave maliciously, while in the Beta Reputation System more time is required to achieve the highest trust value. After negative interactions trust to the user was not significantly degraded due to the previous good experience that the system had with this user. In this experiment a long memory window containing 100 interactions was chosen for referenced models. Our model is based on simpler formalisation and it demonstrates good reaction on the changes in user's behaviour. The time required for regain complete trust of the system is set explicitly by access policies.

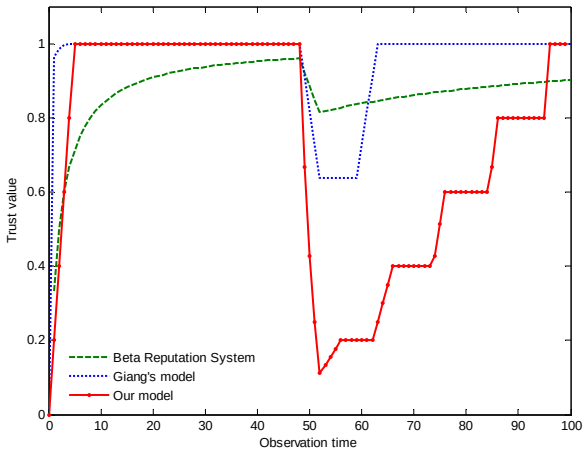
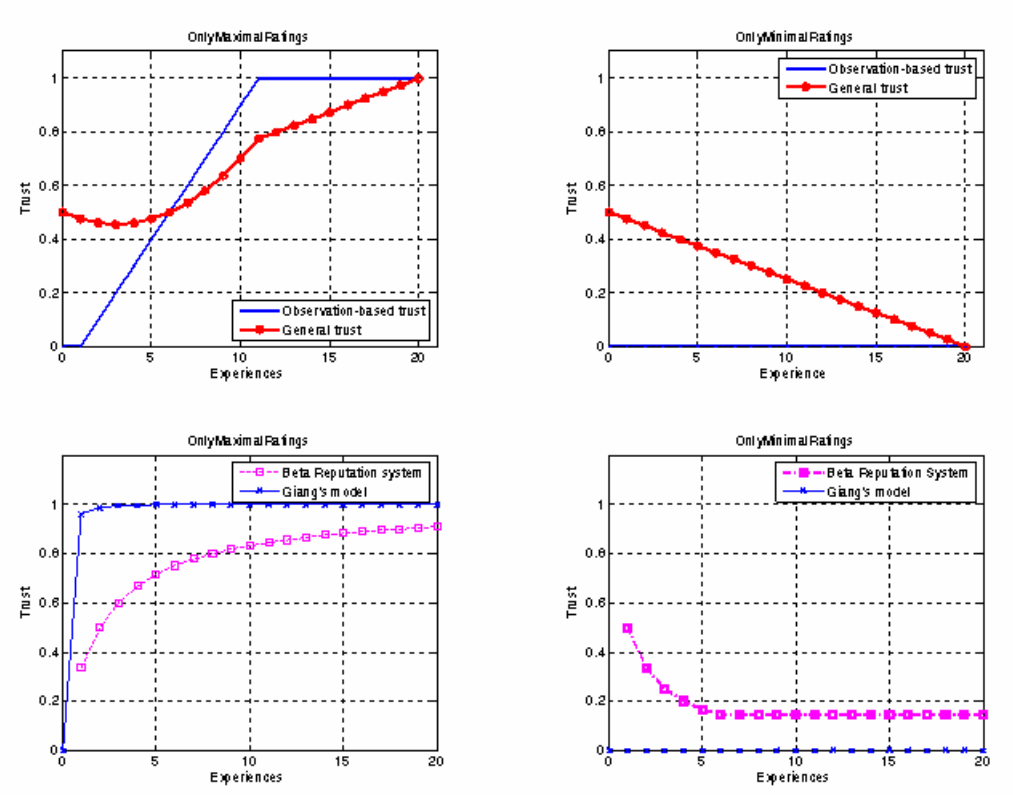


Figure VI.21: Trust development in different models

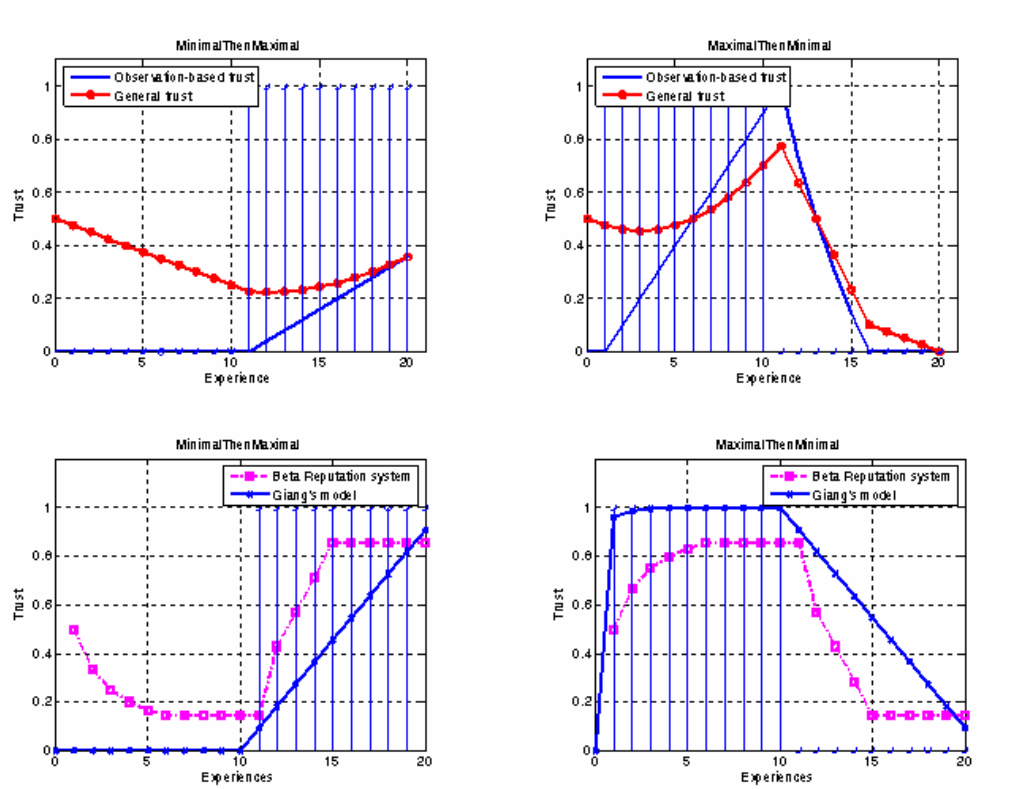
We compare trust dynamics in the beginning of interaction produced by Beta Reputation system, Giang's model and our model. We developed four test scenarios: in the first (Figure VI.22, a) a user demonstrates only good behaviour, in the second (Figure VI.22, b) a user demonstrates only bad behaviour, in the third (Figure VI.23, a) the user starts with bad behaviour and finishes with good behaviour and in the fourth scenario (Figure VI.23, b) a user starts with good behaviour and finishes with negative behaviour. As we analyse the trust development in the situation where the service provider has no sufficient information about a user, the general trust value is defined by the value of reputation of the party that has recommended the user. The recommender's reputation value 0.5 is considered as initial trust.



a b

Figure VI.22: Trust development in Beta reputation System, Giang's model and the proposed model

It can be seen from Figure VI.22 (a) that trust in Beta reputation system rises slowly and approaches asymptotically to its maximum value while the trust value in the Giang's model reaches the maximum value very quickly. In our model observation-based trust reaches the maximum value after the number of positive interactions set by a policy (10 in this experiment), but the general trust value remains less due to the low reputation of the user's recommender. When a user demonstrates only bad behaviour (Figure VI.22, b), our model represents an intermediate solution between the Beta Reputation system, when the user does not loos trust and the Giang's model, when the user does not acquire trust at all. In the proposed trust development the user keep certain trust level caused by the reputation of his recommender until the end of the learning period. After the learning period has finished the user becomes distrusted.



a b

Figure VI.23: Trust development in Beta reputation System, Giang's model and the proposed model, changing of behaviour

When the user changes a behavioural pattern, with the beginning of "good" interactions (Figure VI.23, a) both the Beta reputation system and the Giang's model allow quick achievement of high trust. When the user starts to misbehave (Figure VI.23, b) the Beta reputation system lets him to keep trust due to past positive behaviour. Our model and the Giang's model demonstrate similar results in this scenario.

The main improvement made in this work consists in using a direct and clear relationship between network access policies and trust model parameters. The proposed memory model aims to reduce the space necessary to store long-term user behavioral history in a set of discrete variables, rather than using a time series or description language, as has been proposed in related publications.

VI.8 USER'S TRUST IN A NETWORK

A mobile user collaborates with different entities, which are unknown in advance. Usually a user deals with service providers. Services provided are different by nature. It may be communication services like Internet access, file download services, access to different databases or access to different kinds of information. If mutual authentication is used to access services the user is able to verify if it is dealing with an entity recommended by its home authority or not. In some use cases the information

provided is critical for a user. In other cases it is quite important to associate a certain trust rate with each service provider.

Use case 1: the user has a choice from among various service providers. In the process of candidate selection the user is guided by some well-defined criteria such as provided quality of services, service cost or security services supported. As providers are encouraged to serve users, they may declare their capabilities that differ from reality. The user is motivated to choose fair and reliable service provider for collaboration. Therefore it is desirable to provide the user with a mechanism for the evaluation of a candidate provider's trustworthiness.

Use case 2: the mobile node (user) uses media independent handover services (IEEE 802.21) to optimize the handover procedure. The information service provides the MN information about neighbouring networks and several events related to the link status or other access network parameters. If this information is not correct the MN will not be able to execute fast handover and the active session probably will be affected.

To provide a user a possibility to communicate with only reliable entities a simple trust model inspired by the trust model implemented in eBay [146] may be used. The user gives a trust score ***R(n)*** to each entity *n* that he has collaborated with. If the interaction was successful or the information provided was correct the user increments the score for this service provider. If the user was not satisfied with the services provided, the corresponding score is reduced. All scores are sorted in descending order. If several service providers are accessible from the current user's location, the user is able to select the service provider with the maximum trust score.

The user is satisfied by services provided if they correspond to services advertised and the user is satisfied by information provided if it was correct. For example, the mobile node requests Information Service information about a set of candidate networks. According to response received, one of the candidate networks is a partner of the current serving network and it supports fast re-authentication. The mobile node initiates handover to this network and, in reality, this network does not have roaming agreements with the serving network and fast handover is impossible. A communication session run at the user's terminal is interrupted in and this fact causes "dissatisfaction" by information provided and by the source of this information.

If the user has no information about the candidate service provider or this information is insufficient, the user will communicate only with service providers recommended by his home authority. We suppose that not only the user's home but also serving networks may recommend service providers to a user. A service provider is "known" to a user if interactions were carried out between them in the past. The algorithm of trust evolution for a service provider ***n*** is defined as follows.

```
If service provider (SP) n is unknown and is not
recommended by the home network,
then Do not use its services
if Recommended but unknown SP
   then R(n)=0
If received information was correct
   then R(n)=R(n)+1
        Sort the list of SP by reputation going down;
If received information was not correct
   then R(n)=R(n)-1
        Sort the list of SP by reputation going down;
Define a low threshold of acceptable reputation.
```

The procedure of trust updating is performed after each interaction with each service provider. The user does not keep trust values for a long time for two reasons:

1. Storage resources of the user's terminal are limited.
2. Both the service set and policies of the provider may change over time.

The trust score, used in this context, takes on discrete values. There is no upper bound for this value, it depends on the number of interactions performed between the user and the given service provider. In order to choose a reliable service provider the user defines a *Collaboration threshold CT*. Service providers with a trust score below this collaboration threshold may not be chosen for interaction and therefore should not be kept in the user's memory.

The procedure of candidate service provider selection by its trustworthiness evaluation is performed by the user before authentication with this agent using the following algorithm.

```
1. Extract visible networks from the list of known
   service providers
2. Extract only those that have a reputation value
   greater than a collaboration threshold R(n)>CT
3. Collaborate with the service provider with
   maximum reputation level.
```

To illustrate trust-based service provider selection by the user let us give two examples. The mobile user selects an information provider (IP) to receive handover-related services such as information about neighbouring networks. The mobile node is able to communicate with its home network, but is interested in receiving critical information without delays caused by transmission latency. For this reason the mobile node is motivated to use services provided by one of the networks visible from its current point of attachment.

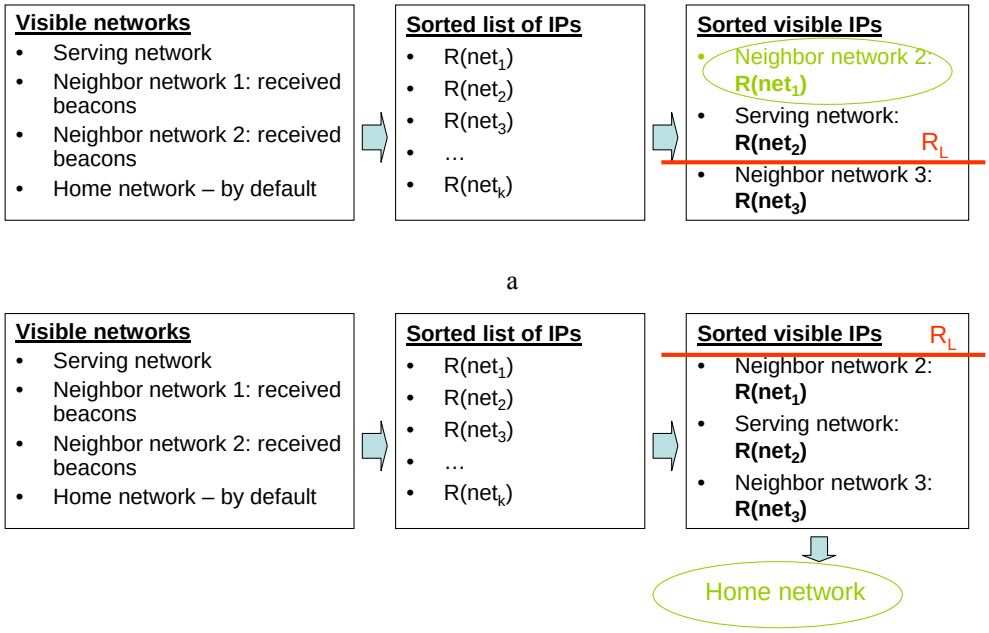


Figure VI.24.Example of candidate service provider trustworthiness verification

We suppose that the user already has some experience with different service providers and he has a list of them sorted by trust score. The user extracts from this list networks visible from its current location and selects the provider with the best reputation (Figure VI.24, a). If there is no visible service provider with a sufficiently high trust score value (Figure VI.24, b), the user selects the home provider.

VI.9 TRUST-BASED ACCESS CONTROL FRAMEWORK IMPLEMENTATION

VI.9.1 System architecture

The proposed trust model is applicable if a service provider can observe, record and analyze the activity of the user. To access the resources each user performs authentication and the network is able to recognize the same agent in different authentication sessions. The entity performing the trust evaluation communicates with the authentication database, the policies database, log files of applications or services, firewalls and intrusion detection systems.

We present a general overview of the components of a trust-based access-control framework and their interaction. Three agents may participate directly or indirectly in each interaction: a service provider, a user and an entity that recommends the user (usually, his identity provider or the home network).

Figure VI.25 depicts the generalized scheme of interaction between the *Trust Evaluation Server* and the *Trust Data Storage*. It also shows the influence of continuous observation of user behaviour on changing access policies. The risk level

is produced from the continuous observation of the behaviour of users and their recommenders.

In addition to the AAA server, each service provider has three new entities, a *Trust Evaluation Server*, a *Trust Data Storage* and an *Observation Agent*. Whenever a user wants to obtain access to services provided by the network, the AAA server authenticates the user. After that it communicates the user identity to the *Trust Evaluation Server* that performs a trust evaluation of the user based on information about his past behaviour and the current risk level in the environment taken from the *Trust Data Storage*. If the user is considered to be trusted, he is given access to a set of services according to the assigned trust level and the presence of a recommendation from a party trusted by the service provider. If the user is distrusted, service access is denied him. The *Observation Agent* records the behaviour of the user during the session held and transmits the data collected to the *Trust Data Storage*. The *database of users* contains information about the past behaviour of visitors and the database of operators contains information about the past behaviour of partners, whose subscribers have been served by this network. If the network provides services for a fee, other operators may be responsible for transfer of user payments for services accessed. In such a scenario, the partner will be considered a "bad" partner if it delays payments or does not transfer them. The trust values for the user and for his home network are calculated using information from corresponding databases.

When an agent has the role of *recommender* in a current interaction, it needs to decide whether it issues a recommendation for a user or not. To make this decision, the recommender calculates the trust value of the user based only on the user's past behaviour recorded in the *Subscribers data base* with the help of the *Trust Evaluation Server*.

On the user's side, the following elements are involved in a provider's reputation evaluation: the *Service Providers* database, the *Observation Agent* and the *Trust Calculation Engine*. The database of service providers contains identifiers of service providers, sorted by the reputation value. The functionality of the *Observation Agent* is limited to measuring a certain parameter and comparing it with the required value. For example, the declared and the provided QoS or the price paid for services used may serve as a basis for reputation construction. Operations performed by the *Trust Evaluation Engine* on the user's part, are limited to a simple comparison of the measured and desired values of evaluation criteria. Recommendations received from the home or other trusted authority may be stored locally and may be presented to the service provider with which the user wishes to collaborate.

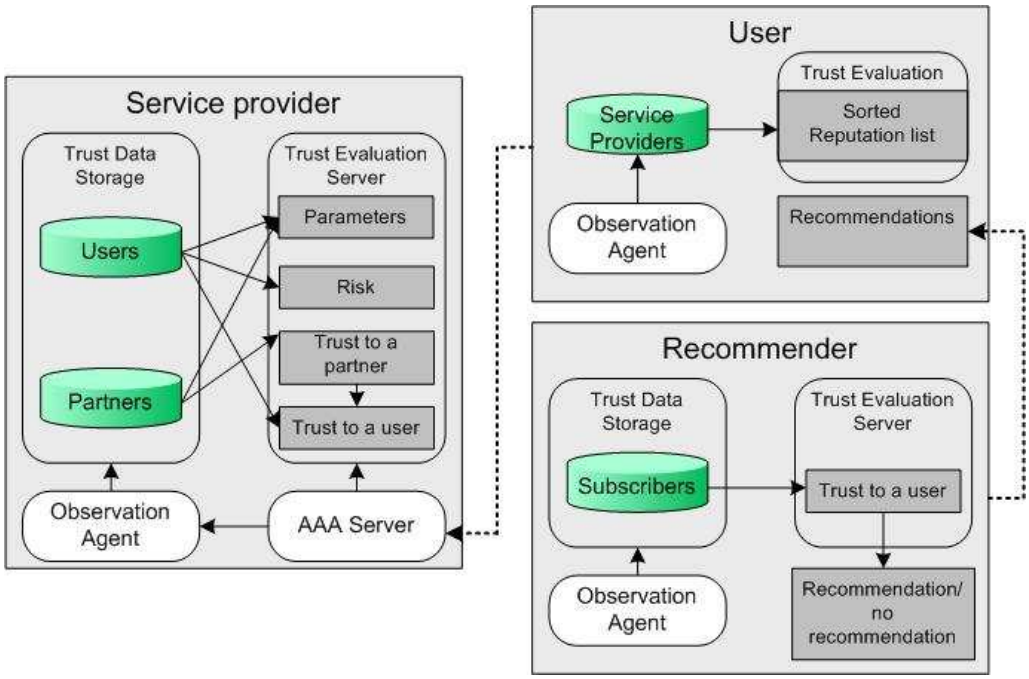


Figure VI.25.General architecture of the trust-based access control framework

The proposed access control system relies on trust calculation and consists of static and dynamic components. We call a component static if it is not changed in an autonomous manner. The service sets and corresponding trust levels are static components because changes are performed by the system administrator. We consider that the proposed system is not able to add or remove services automatically.

Databases containing information related to users and to partner service providers change over time, causing development of the risk value observed, and as a consequence service access policies.

In the digital world it is very simple for a user to change its identity if it is an e-mail account. We assume that the proposed trust model is designed for access networks that implement strong authentication and identity or role-based access control. Anonymous access is the current practice for peer-to-peer or overlay networks or for web-services. The concept of user identity, adopted in this work, is that defined in the 3GPP. Each user has a pair of identifiers, public identity and private identity. The main requirement is that the identity must allow users' traceability.

VI.9.2 A use-case scenario

This section describes the authentication and access control framework by putting together the components introduced earlier. A mobile user evaluates trustworthiness of each available access network using an approach presented in Section VI.8. In Section V.2 we provided an overview of the protocol for fast authentication in an inter-domain handover scenario. The optimized scheme for authentication tickets distribution is described in Section V.3. Authentication results in giving a user authorization to access a network, but usually an access network should decide what are the authenticated user's access rights and privileges. After user identification the access network evaluates trust in this user by the method introduced in Chapter VI, parallel to authentication process execution.

Figure VI.26 demonstrates that a user can associate with access networks that may be based on different technologies and that are managed by different authorities. Let the user John be subscribed to the operator of network A as well as to the operator of network B. Operator C is a roaming partner of operator A and serves subscribers of its partners. Network D is a network managed by a non-profit operator, which may be a school or enterprise authority. Network D has no agreements with its neighbours. The mobile user John does not want to expose his identity to a non-home authority, which is why he is perceived as *John@A* (or *John@B*) by his identity providers, and as for example *i_am_away@A* and *i_am_away@B*) by non-home networks. John may use the "external" identity for network D or have a specific identity for it.

We assume that there are other access networks coexisting in the same geographical area, but the user has considered them unreliable, which is why he does not take their advertisements into consideration.

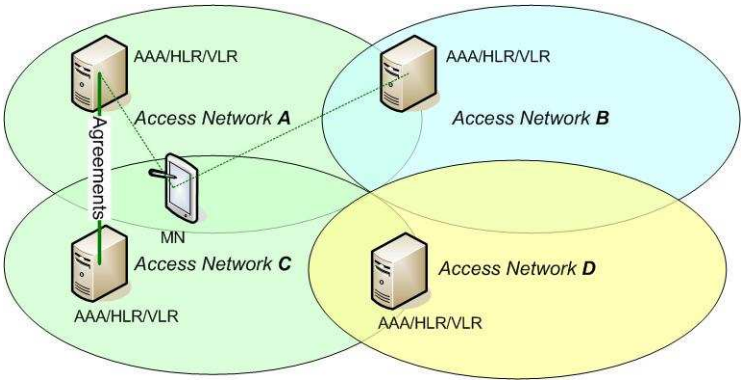


Figure VI.26: Handover scenario: four access networks managed by different authorities

Modern mobility management and fast authentication protocols allow fast transition to be performed only between partner networks A and C, which can exchange user-related security context in a secure manner. When associating with another network, the user must execute full authentication exchange.

If access networks are FAP-enabled, the user in the described use-case scenario can perform fast authentication with any network in a region. Each network issuing fast authentication tickets may issue them for itself. In that way the user may handover from network A to network B using a ticket issued by B and, being either in network B or network C, handover to network D using its own ticket.

If the user has demonstrated malicious or suspect behaviour in network C, which is a roaming partner of his home network, service provider C may restrict the user's authorizations despite the presence of roaming and service level agreements with the user's home network.

VI.9.3 Authentication and authorization

For illustration purposes we use IEEE 802.11 here to demonstrate the proposed authentication and authorization framework. To address fast handover requirements, the procedure of authorization rights definition must not increase authentication latency. In this regard, access policies enforcement should be made in parallel with user's authentication and session keys negotiation.

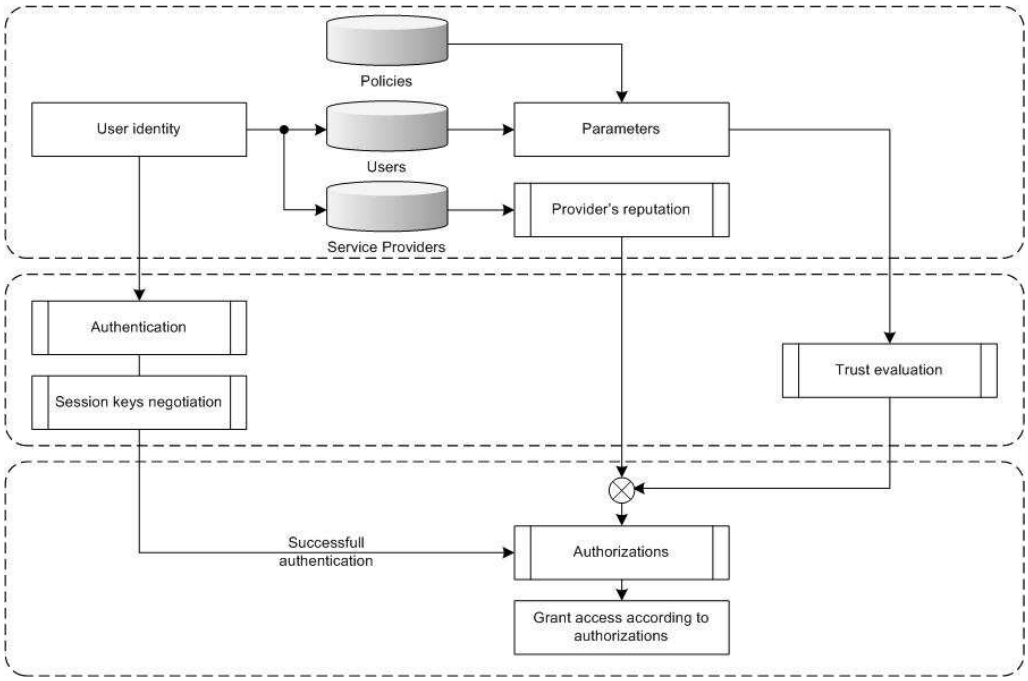


Figure VI.27: Process of user authentication and authorization in a visited network

After receiving the user's identity response, a RADIUS server typically searches for the corresponding user in its local database. If interaction history-related data is kept along with a username, the authentication server has all the information necessary for trust value evaluation at the end of the authentication process. Thus the calculation of the trust value may be started at the same time as the session keys negotiation process. Figure VI.27 shows the succession of actions performed by the authentication server within an access network in order to authenticate a user and to determine his authorization in the session to come. The first block of actions includes user identification and searching for information related to his past visits. If the user is not found, at least its recommender (identity provider) must be known to the serving network. The second block of actions includes fast authentication and the subsequent block serves to determine user's authorizations, which are based on the trust value of the user, his recommender or both, as it is described in Section VI.5.

VI.10 CHAPTER SUMMARY

In different works trust is represented as the combination of the probability of future acceptable behaviour of a partner, competence, disposition, benefits, costs, witness, recommendation and past experience. In this work we design a trust model for access control in networks that provide services. We assume that service providers are motivated to serve users because of the profit received from interactions. Users need to have high quality access to and use of their preferred services. As the service provider does not know potential users it should make several hypotheses concerning user behaviour. It is natural for a service provider to accept users that are unknown but recommended by a business partner. It is also natural to stop serving a user, even recommended by a partner, if the former displays malicious behaviour. The service

provider can reserve the highest QoS for privileged clients that have demonstrated good behaviour (satisfy network policies) for a certain period of time.

To deal with an unknown user, recommendations from a trusted party have decisive influence on the decision making concerning the trustworthiness of this user. Over time personal experience becomes more important than the recommendation, because a recommender may consider the user to be a good one yet not be aware of his behaviour.

The proposed trust model may be implemented to improve access control in open environments such as wireless networks of Internet service providers that serve a large number of users. This model is also suitable for peer-to-peer environments such as grids or file-sharing systems. The generalized formalization of notions of trust, behaviour and risk allows the model to be suitable for various deployment scenarios. In such scenarios each peer is at one and the same time both a user and a service provider.

Trust formalization is different for the service provider and the user. This results from the fact that trust of the service provider in the user must be refined to allow or deny different kinds of actions within the same interaction, while the user has only a binary choice, to interact or not to interact with a service provider.

In this work we consider the aspect of trust development over time rather than different aspects of trust propagation. The main improvement made in this work consists of using a direct and clear relationship between network access policies and trust model parameters. The proposed memory model aims to reduce the space necessary to store long-term user behavioural history in a set of discrete variables, rather than using a time series or description language, as has been proposed in related publications. A linear trust model provides the best performance when compared with non-linear models described in the literature.

The main point of originality of the proposed model is the use of different sources of trust, the possibility of dynamic adaptation to the changing environment, and the ability to work with user history over a long timeframe.

Conclusions and perspectives

VI.11 CONCLUSIONS

In this thesis we have examined the challenges related to providing secure ubiquitous mobility in the heterogeneous universe, from the perspectives of both users and the access network operators. The heterogeneity of modern wireless access technologies sets particular requirements for mobility management and security mechanisms design. On the one hand, such mechanisms must be flexible, self-organizing and independent of the underlying media technology. On the other hand, the absence of physical access limitations in wireless networks and their openness create a need for strong confidentiality and integrity protection at the link-layer. For its part, mobility brings up new challenges for providing security.

In Chapter IV we demonstrated our vision of future wireless networks; then we identified key requirements to provide ubiquitous, secure mobility. Trust becomes a central concept for security mechanisms design. Contributions presented in this thesis help to solve the following problems:

- ✓ **Roaming region extension**: The network managed by his home provider or its partners limits the roaming region of a user. In our trust model, the user's roaming region may be defined not only based on the contractual agreements between service providers but also on trust relationships between a specific user and network access providers.
- ✓ **Secure network selection**: Users require access to their preferred services and to enjoy high quality use. If numerous access networks are available, users wish to select the most trusted network access provider. To achieve this, we introduce a rating-based scheme for candidate access network selection. In this work we did not consider how a user learns about network-related information.
- ✓ **Dynamic trust establishment between previously unknown parties**: In our Fast re-Authentication Protocol, we eliminate the need for communication between the target and the user's home networks in order to establish trust between a visited network access provider and a mobile user. We propose a trust delegation mechanism that is based on authentication tickets.
- ✓ **Heterogeneity of credentials**: the authentication ticket used in the proposed Fast re-Authentication Protocol contains information that is based on the user's credentials and the result of the previous authentication, but is independent of the credentials type and the authentication method used.
- ✓ **Fast authentication in inter-domain handover**: Decreasing inter-domain handover authentication time is achieved by Fast re-Authentication Protocol (FAP). FAP localizes the authentication process, eliminates the need for heavy management of user credentials and minimizes communication

between different administrative domains. The method does not require centralized data storage or topology sharing between different service providers. FAP allows mutual generation of key material, which serves to produce session encryption keys. The protocol consists of ticket acquisition and authentication phases. Signalling optimization combines location update to the home network or a broker with the request for credentials. To minimize the number of authentication tickets sent to each subscriber, we propose the use of a neighbour table, which is maintained by an authentication server on each network. When the client requests a ticket, the server generates tickets for only the networks contained in the line of the neighbour table corresponding to the current location of the user. This method decreases the number of tickets sent, and consequently the overhead and delay of the protocol's ticket acquisition phase.

Numerical results obtained from experiments on a test-bed and a series of simulations show that the proposed method enhances inter-domain handover parameters such as authentication latency and signalling cost.

- ✓ **Access control in open environments:** Networks dealing with a great number of potentially malicious users are capable of automatically adapting access policy to the circumstances. The history of previous interactions is formalized in a lightweight manner; a network server does not store a large amount of user-related data; in the proposed model, access control policies are adaptable for changes in users' and partners' behaviour. All parameters of this model are directly defined by policies set in a natural language and are easy to understand. The ability of the proposed trust-based access control model to mitigate attacks is confirmed by a series of simulations. The proposed scheme for trust-based access control may be implemented for decentralized architectures such as peer-to-peer or grid networks as well as for centralized architectures.

VI.12 RESEARCH PERSPECTIVES

Heterogeneity of architectures, access technologies and protocols is a source of enormous possibilities and great challenges. Contributions made in this dissertation still need further study, in particular in fields of validation and operational implementation. Regarding this work we define the following short-term research directions:

- ✓ The compound user authentication approach, introduced in Chapter 5, requires more detailed performance analysis with different scenarios of IP address acquisition and service access organization. The resistance to mentioned attacks also should be demonstrated. Further simulations and experiments on a test-bed can be carried out to achieve this goal.
- ✓ The prototype implementation of the EAP-FAP can be further developed by including the ticket acquisition phase. More experiments with different attack scenarios should be carried out to prove the protocol's resistance and to study its performance under attacks.

- ✓ For authentication protocols proposed in this thesis there is a need to formally validate their security properties. Formal analysis can detect potentially vulnerable states in protocol's operation.
- ✓ We have proposed a high-level, generalized mechanism for trust-based access control. Difficulties in implementing this mechanism consist of the need for statistical analysis of user behavioural patterns with various access control policies.

Mobility related issues mentioned in this thesis require further investigation and we identify the correspondent research directions as follows:

- ✓ There is a need to manage trust relationships between service providers in a dynamic manner. With the use of the proposed trust model these relationships can depend on users' behaviour and an honesty of recommendations.
- ✓ It would be desirable to extend the context of the trust model and to determine explicitly what user behaviour is considered to be "positive" and what is considered to be "negative". To achieve this goal the functionality of user activity observation tools needs to be studied.
- ✓ The proposed trust model may be implemented to improve access control in open environments such as wireless networks of Internet service providers that serve a large number of users. The generalized formalization of notions of trust, behaviour and risk allows the model to be suitable for various deployment scenarios. To make this model suitable for peer-to-peer environment feedback definition should be incorporated into the model.
- ✓ Long-term fast authentication credentials that are issued for reliable users and allow them to handover to a credentials issuer or its partner even from a non-partner network should be introduced.
- ✓ Our contributions do not yet address accounting and billing a mobile user in a network owned by a non-home operator. Using existing accounting mechanisms, a user is unable to verify or check the information about services used that the visited provider makes available to the user's home provider. Off-line accounting schemes need to be designed. If a network provides chargeable services, the payment factor must be accounted within a trust-based access control mechanism.
- ✓ In this thesis we have defined the problem of secure network discovery and handover preparation. A possible extension of this work would be to define media-independent, secure mechanisms for pre-authentication information exchange between the mobile user and a set of candidate networks. This work could be included in the IEEE 802.21 Security Task Group handover signalling optimization and Media Independent Handover protocol security framework.
- ✓ Another issue needed to be addressed covers security policies mapping and security level maintenance for interacting entities.

ANNEX A

OPTIMAL TICKET DISTRIBUTION: SIMULATION MODEL DESCRIPTION

To analyze the protocol performance, a model was created using OmNet++ [147].OMNeT++ is a public-source, component-based, modular and open-architecture simulation environment. Components (modules) are programmed in C++, and then assembled into larger components and models using a high-level language (NED).

Each simulation was held in a roaming region covered by 16 access networks. Each network operator can have roaming agreements not only with neighbouring networks, but also with other networks in a region being studied as shown in Figure A.1. Dark points correspond to analyzed network operators and grey points designate its roaming partners in the region studied. All operators have an equal number of subscribers.

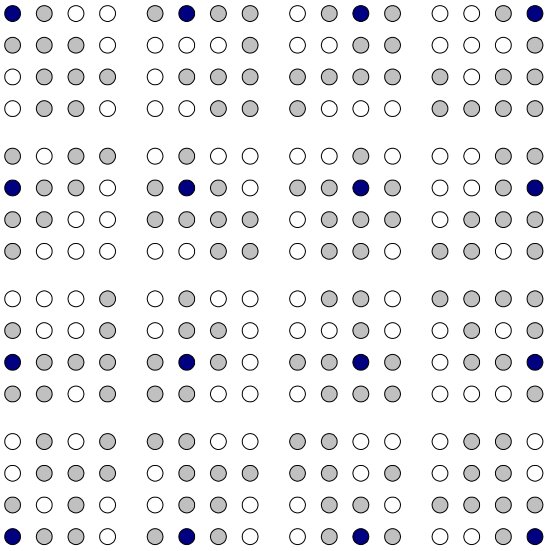


Figure A.1: Simulation setup: roaming agreements of each operator

At the start of the simulation, users are distributed uniformly across all partners of their home network, and each network has an empty neighbour table. Each client chooses the roaming destination randomly with uniform probability. As intra-domain authentication is beyond the scope of this study, our simulation model does not include re-authentication in cell handovers. We have defined three types of user mobility, low, medium and high. Each type of mobility is characterized by the time interval between two consequent inter-domain handovers.

The duration of each simulation was 24 simulated hours. By the end of simulation, the neighbour table has been created for any mobility type, and all authentications are executed in the proactive mode.

ANNEX B

VALIDATION OF TRUST-BASED ACCESS CONTROL MODEL

B.1 Simulation model description

We have evaluated the performance of the proposed trust-based access control method via a series of simulations. These simulations are realized using OMNeT++ discrete event simulator.

The studied model consists of three parts:

1. User mobility modelling;
2. Implementation of the trust model on the server side;
3. Implementation of the trust-based access control.

A simulated service provider's network supports trust-based access control. Users served by this network are mobile; they are characterized by different visit frequencies and different behaviour ("good" and "malicious"). We did not consider authentication in this model.

The aim of the simulation was to study the evolution of trust to each user in a network as a function of the behaviour of all visitors. The simulation model consists of a network and a set of its clients. At each moment a client may be either attached to this network (or be requesting access) or be away.

User's behaviour is described by the following parameters:

1. Average duration of the visit (seconds) and frequency of visits (1/second);
2. Probability of "bad" behaviour; Note that we do not define what "bad" behaviour means.

The access network retains user-related data in a following structure:

```
struct client_t{
              int name; //client index()
              int pos; //number of positive experiences
              double neg; //number of negative experiences
              simtime_t lastVis; //time of the last visit
              simtime_t validUntil; //limit of trust validity
              double trustVal;
              int num_fails; // the number of times the client
was distrusted
              bool good; // false if the client is currently
distrusted
              bool never; //true if must not be forgiven
       };
```

We have defined the following trust-related policies:

```
int beg_trust; //number of positive experiences to gain trust
int stop_trust; //number of negative experiences t lose trust
int num_visit; //number of consequent positive experiences to
regain trust level
int forgive_t; //time after which a "bad" user may be forgiven
double max_trust[5];
```

After a client has disconnected, the trust calculation module updates client-related information and access policies. Trust values for each client and the corresponding global negative behaviour rate are retained for further analysis. These values are presented in a vector form as functions of time.

```
cOutVector Trust;// trust evolution for all clients
cDoubleHistogram trustStats;//statistics for trust evolution
(average, min, max etc.)
cDoubleHistogram negStats; // statistics for negative rate
evolution (average, min, max etc.)
cOutVector Neg; //negative rate calculated over all clients
```

B.2 Simulation scenarios

As it was mentioned in Section VI.2.2 a service provider deals with four types of clients, good, early bad, random bad and strategic bad. A strategic bad user may increase the frequency of visits when performing an attack to attain his goal more quickly and before the network adopts countermeasures. Since the impact of each user's behaviour is not isolated from others, in our simulations we considered different scenarios of bad user's activity.

Scenario 0: 50% of clients are not fair and they perform malicious actions with the probability 0.5. Each malicious client may start to attack the network from the very beginning of interactions as well as after a certain number of visits defined by the parameter `myTrust.begin`.

In Scenario 1 (Figure B.1 a) the system is attacked at a particular moment and the number of attackers increases gradually. In Scenario 2 (Figure B.1 b) all involved attackers stay active for a certain period of time. The situation in which users starting to misbehave keep up malicious activity is shown in Scenario 3 (Figure B.1 c). Finally, Scenario 4 (Figure B.1 d) depicts a state of permanent "war" with a short period of armistice.

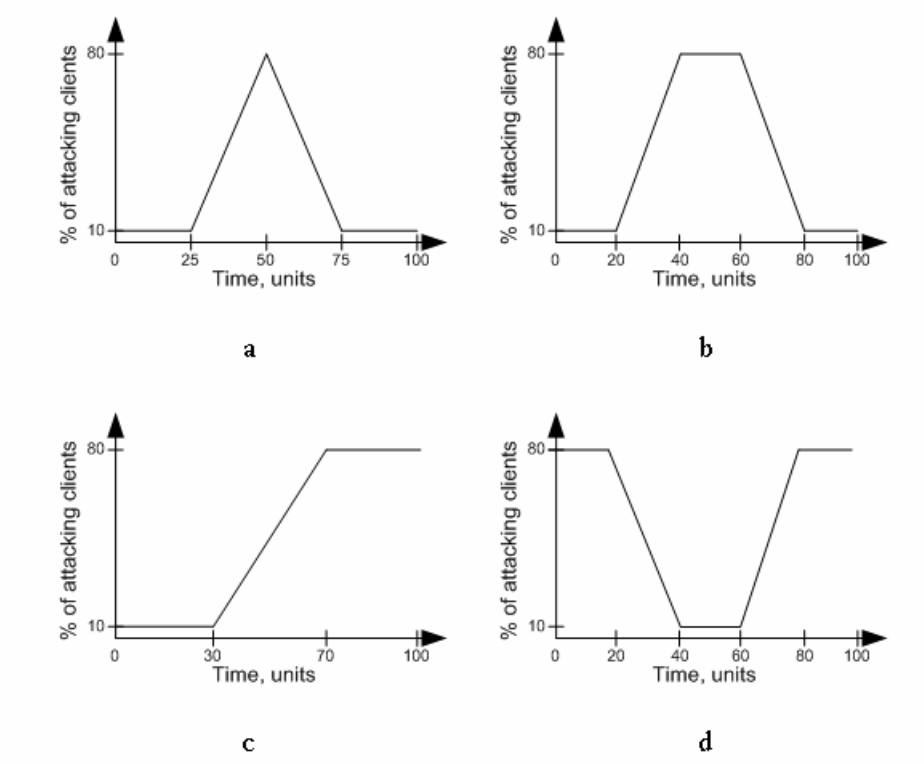


Figure B.1: Attack scenarios

In scenarios 1 – 4 the maximum number of attacking clients is not more then 80% of the total number of clients.

B.3 Simulation setup

To estimate effectiveness of the implementation of adaptive trust-based access policies we ran experiments of two types. In the first series of experiments static policies were used. These policies were defined manually at the beginning of the simulation. In the second series of experiments the initial values for trust policies were the same as in the first series, but a mechanism for policy update described in Section VI.6.5 was implemented.

Figure B.3 shows dialog boxes to set up simulation parameters. The first dialog box allows a choice between the static and adaptive trust policies. The parameter `myTrust.begin` defines after how many visits a malicious client starts to misbehave. `myTrust.model` specifies what attack scenario will be used in the simulation. Whether a client changes or not the frequency of visits after starting an attack is determined by the value of a logical parameter `myTrust.active`. The two last depicted dialog boxes are to set up initial values for trust policies. For illustrative purposes, only the number of positive experiences to gain trust and the number of negative experiences to lose trust were chosen to be flexible in this series of simulations. All other policies were fixed.

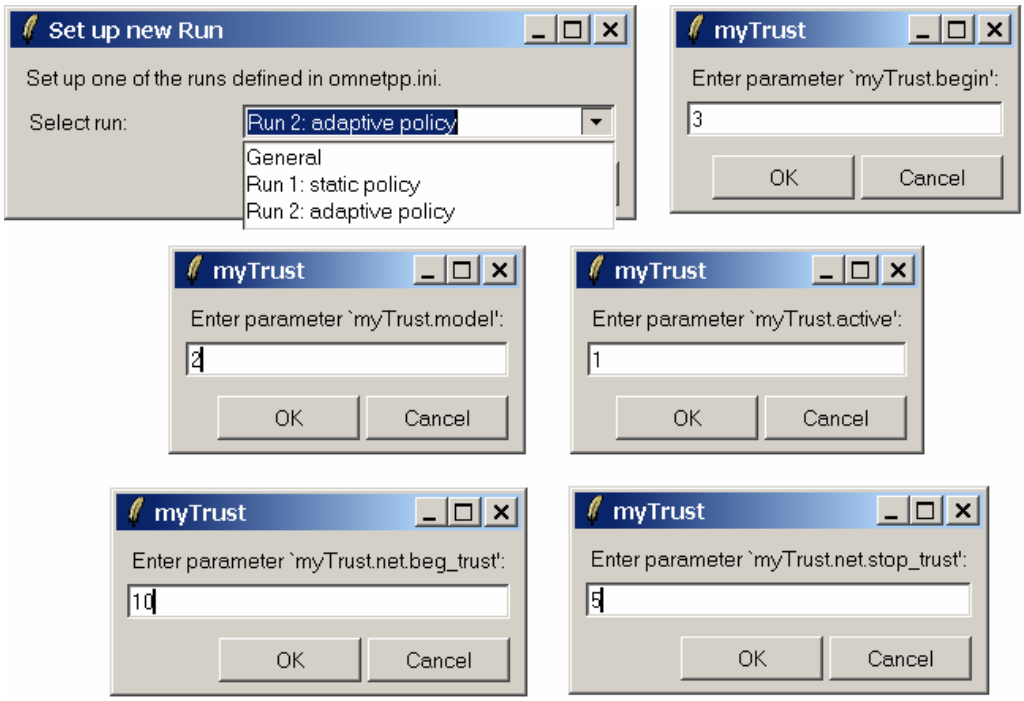


Figure B.3: Setting up simulation parameters

We simulated a service provider's network interactions with 200 clients. Initial values for `myTrust.beg_trust` and `myTrust.stop_trust` parameters were set to 10 and 5 correspondingly. To update policies only one negative rate threshold was set. If negative rate exceeds the value 0.1 then the value of `myTrust.beg_trust` is increased by 5 and the value if `myTrust.stop_trust` is decreased by 4.

Each experiment lasts 100 simulated days. There were five trust levels defined ([0, 0.2) [0.2, 0.37) [0.37, 0.63) [0.63, 0.8) [0.8, 1]). The forgiving period for a distrusted client was set to 10 days and the number of visits to regain trust level was equal to the `myTrust.beg_trust` parameter.

An example of a simulation run is shown in Figure B.4. Distrusted users are marked in red color. A client requesting network access is emphasized.

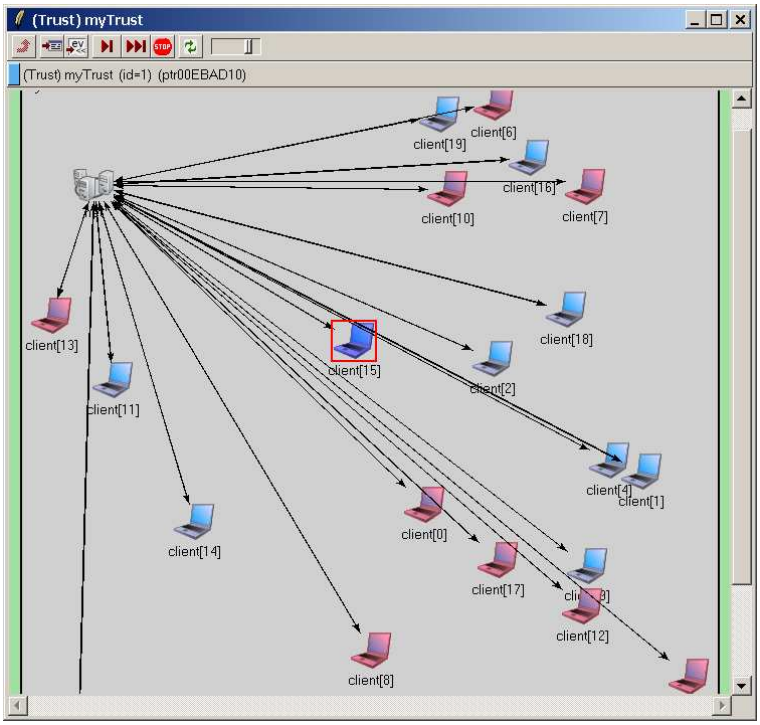


Figure B.4: Interaction between service provider and clients

B.4 Policies adaptation in different attack scenarios: simulation results

Figures C.5 – C.9 depict negative rate development under attack scenarios described in Section C.2. Blue dashed lines correspond to the observed rate of negative behaviour for a system implementing static access policies and red lines correspond to the rate of negative behaviour observed in the system implementing adaptive access policies. In the first scenario there is no need to update access policies due to the constant and low attacker activity. It can be seen that adapting policy helps to mitigate an attack, as compared with the model with static policies in scenarios 1 - 3. On the other hand, in the "war" scenario 4 adapting a more nuanced policy is needed to protect the network. .

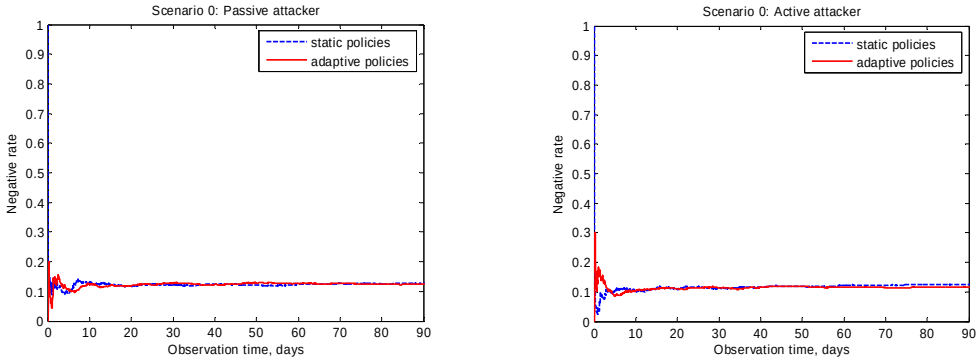


Figure B.5: Simulation results for Scenario 0

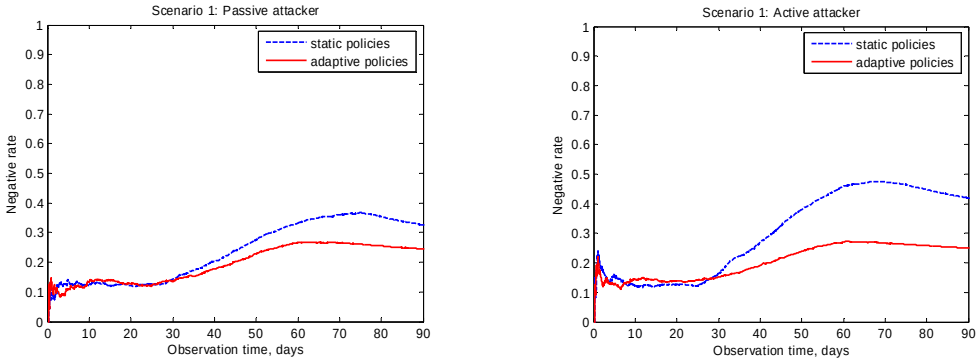


Figure B.6: Simulation results for Scenario 1

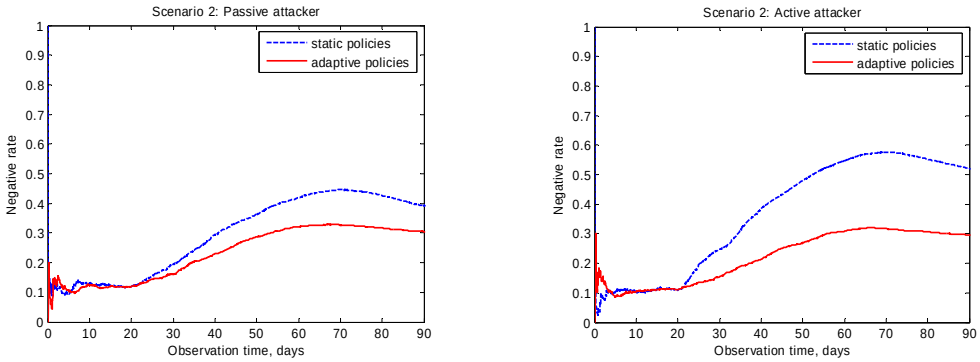


Figure B.7: Simulation results for Scenario 2

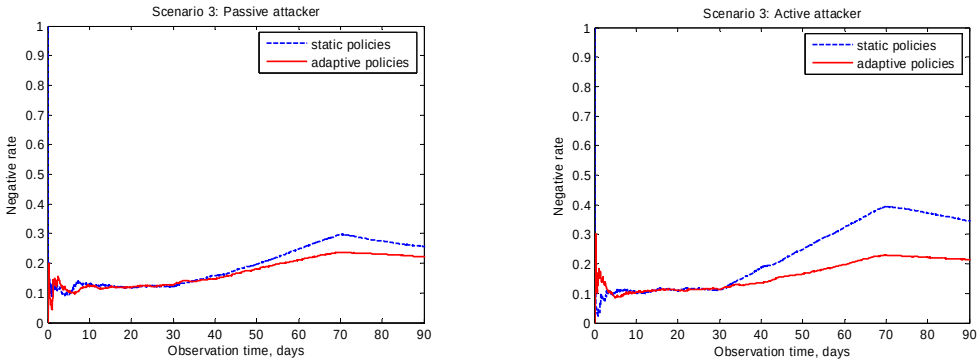


Figure B.8: Simulation results for Scenario 3

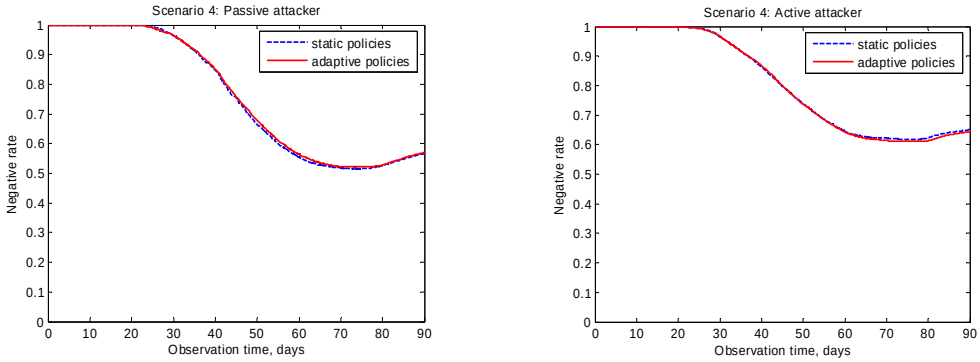


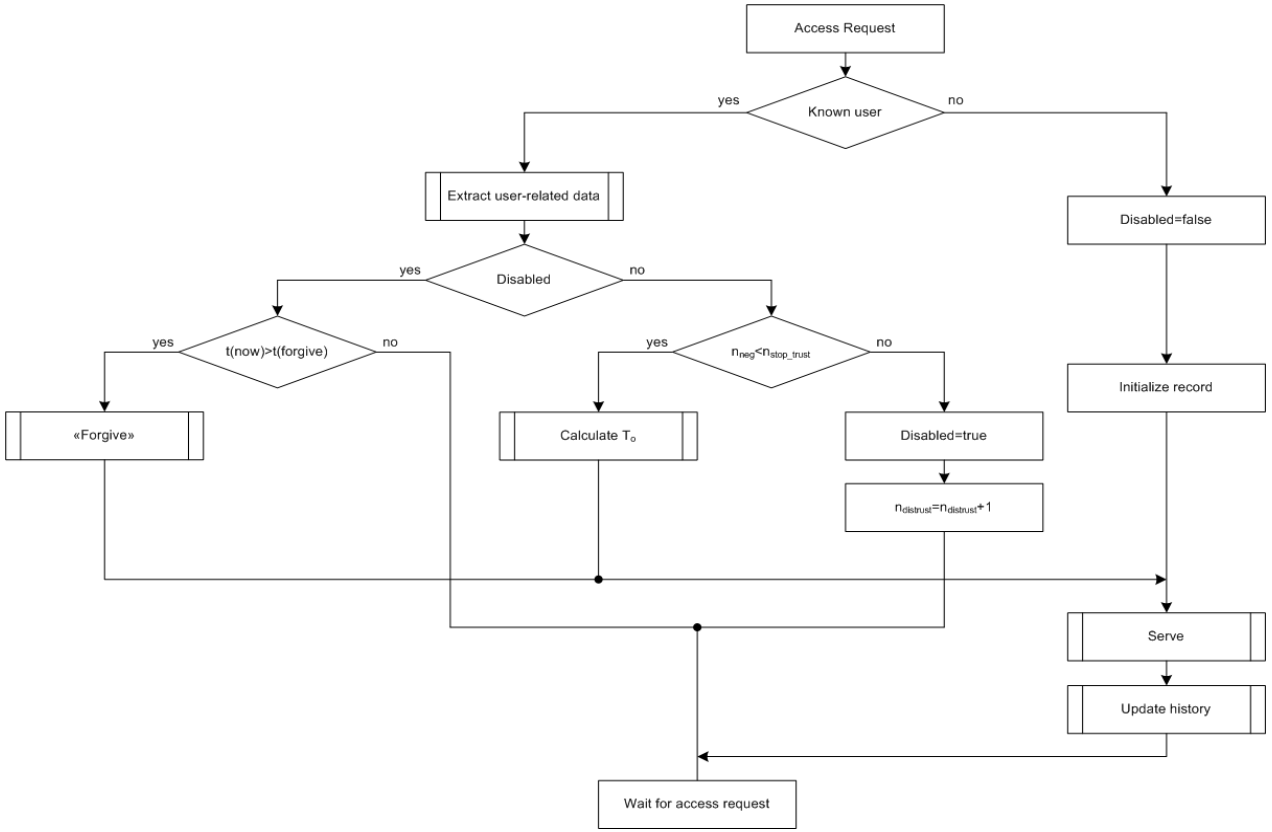
Figure B.9: Simulation results for Scenario 4

Table B.1 summarizes results of simulations for each attack scenario.

Table B.1: Comparison of static and adaptive policies-based models

	Passive attacker			Active attacker		
	Static policies	Adaptive policies	Δ	Static policies	Adaptive policies	Δ
Scenario 0	0.14	0.13	0.01	0.13	0.12	0.01
Scenario 1	0.37	0.27	0.1	0.48	0.27	0.21
Scenario 2	0.45	0.33	0.12	0.58	0.32	0.26
Scenario 3	0.33	0.24	0.09	0.40	0.30	0.10
Scenario 4	1	1	0	1	1	0

B.5 Algorithm for observation-based trust calculation



ANNEX C

CONTRIBUTIONS TO IEEE 802.21 (MEDIA INDEPENDENT HANDOVER) SECURITY TASK GROUP

C.1 Problem statement for MIH authentication and analysis of possible solutions

In this analysis of the applicability of pre-authentication schemes for inter-domain handover we consider two scenarios, handover between networks of the same media technology and handover between networks of different media technologies.

C.1.1 Use cases and Requirements

Scenario 1. A mobile device moves between two networks of the same media type deployed in different administrative domains.

The mobile node is attached to the serving authenticator and it has chosen a set of candidate authenticators. To attach to a candidate network, the mobile node must perform authentication with an authentication server via an authenticator deployed either at the link-layer or at the network-layer. The target authentication server may be located in the same local network as the target authenticator (local AS) or in another network (remote AS).

In this case two situations are possible:

1. There are roaming agreements between the serving and the target networks. The target authenticator and the target authentication server trust information forwarded or signed by an entity belonging to the serving network. According to these roaming agreements, trust relations should be established between the Serving Authenticator (SA) and the Target Authenticator (TA).
2. There are no roaming agreements between the serving and the target network. There are roaming agreements between one of the mobile node's home networks and the target network. This situation is possible when the mobile node has subscriptions in multiple networks.

The target authenticator and the authentication server trust information forwarded or signed by an entity belonging to the mobile node's home network.

To provide pre-authentication in this use case the following requirements must be addressed:

1. A way by which information about the target network is provided to the MN should be defined.
2. Protocols used for communication between entities of the MN's home network and entities of the target network.

Scenario 2. A mobile device moves between two networks of different media types and deployed in different administrative domains, e.g. 802.16 and 802.11.

It is important to note that security signalling for handover between different media types is not equal in different directions. In this use-case two assumptions apply:

1. A mobile device moves from either 802.16 or 3GPP to 802.11 access network.
2. A mobile device has 802.11 network interface activated.

To perform the proposed solution the target authenticator must support pre-authentication.

C.1.2 Pre-authentication applicability to inter-domain authentication

To optimize the security signalling for the inter-domain handover EAP pre-authentication has been proposed [,148]. Two modes of EAP pre-authentication are defined: direct pre-authentication and indirect pre-authentication. This section provides a study of the applicability of this approach for defined use cases.

Issue 1.1 is related to management of pre-authenticated mobile nodes. The mobile node normally does not know the lifetime of a pre-authentication session.

Issue 1.2 concerns resource consumption: the mobile node makes use of different MIIS (Media-Independent Authentication Services) and pre-authenticates with every potential candidate authenticator while the pre-authentication result is already kept by the AAA server. When the mobile node performs pre-authentication with all candidate networks, with growing number of mobile users and the high density of access networks in a geographical region the traffic overhead increases very significantly on authenticators and authentication servers.

Direct pre-authentication

It is assumed that the target authenticator accepts direct pre-authentication. In this case the serving authenticator is not involved in the pre-authentication signalling. Pre-authentication is performed either at L2 or at L3 depending on which layer the target authenticator is deployed.

To perform direct pre-authentication, the mobile node requires support from 802.21 only in terms of providing information about the target authenticator address and supported pre-authentication mode.

Within this scenario the mobile node can send to the target authenticator multiple, consecutive pre-authentication requests. After successful authentication the AAA server sets a lifetime for the pre-authentication session. The mobile node is not aware of the session expiration time (Issue 1.1) that is why the mobile node can perform the pre-authentication with a valid non-expired, pre-authenticated state. This situation is possible if

1. The mobile node did not handover after pre-authentication and aims to extend the pre-authentication lifetime;

2. The mobile node handovers to another candidate authenticator and a new set of candidate authenticators may include the same authenticators as the previous set of candidate authenticators.

Issue 1.3. The authentication server performs a new pre-authentication upon receiving a new pre-authentication request from a pre-authenticated mobile node. In this way direct pre-authentication makes the target authenticator vulnerable to DoS attacks: if the mobile node is an attacker it can send numerous pre-authentication requests to the same TA.

An attacker can observe pre-authentication requests arriving in clear and after that it can impersonate a valid user by sending pre-authentication requests on its behalf. As the attacker is not aware of the mobile node's credentials, the pre-authentication will fail and the valid mobile node's state will change to "unauthorized". When the valid mobile node will try to perform a fast authentication using the MSK generated as a result of the pre-authentication, it will not have a corresponding authorization and will execute the full authentication exchange with the target authenticator.

To eliminate the possibility of described attacks, the indirect authentication proposed in [86] can be implemented.

Indirect pre-authentication

The target authenticator may process pre-authentication requests only from authorized nodes. If there are roaming agreements between the serving and the target networks, the serving authenticator can forward pre-authentication requests from the mobile node to one or many candidate authenticators. In this case infrastructure knowledge sharing and trust relations establishment are needed between authenticators belonging to different administrative domains. The following issues come to light:

Issue 1.4 is related to the level of authenticator deployment. For example, the serving authenticator may be implemented at L2 and the target authenticator – at L3 or vice versa. The pre-authentication problem statement draft [86] specifies that the indirect pre-authentication signalling is performed over L3.

The L2 authenticator generally not authorized to communicate with entities in other subnet or administrative domains. The deployment details of the L2 authenticator or additional entities and relations with them should be defined to assure the possibility of communication between the serving and the target authenticators for pre-authentication reasons. There are two possibilities to solve this problem:

1. Add functionality to the L2 authenticator. This solution leads to increasing the cost of deployment of the authenticator.
2. Locate the forwarding function at the access router. In this case the transport and interfaces between the authenticator and the access router should be specified.

Issue 1.5: If the mobile node performs indirect pre-authentication via the serving authenticator, the latter must distinguish the pre-authentication requests to be processed from pre-authentication requests to be forwarded to the target authenticator.

Issue 1.6: The functionality of the serving authenticator for indirect authentication is only defined for 802.11.

C.1.3 Applicability to roaming agreements use cases

Roaming agreements specify which network entities may communicate with each other.

Within use case 1 roaming agreements are established between the serving and the target networks and either direct (if supported by the target authenticator) or indirect (if supported by both the target and the serving authenticator) pre-authentication can be performed.

Within use case 2 roaming agreements are established between the serving and the mobile node's home networks. Pre-authentication is possible in direct mode, if this mode is supported by the target authenticator.

Issue 1.7. Indirect pre-authentication that involves the serving authenticator is not possible in this case even if the target authenticator supports this pre-authentication mode.

The AAA protocols (RADIUS and Diameter) can operate in proxy mode forwarding authentication messages from the mobile node to the destination AAA server. The mobile node can perform pre-authentication via its home network. If such candidate networks were chosen, indirect authentication can be accomplished by splitting pre-authentication signalling into MN – HA and HA – TA signalling. HA defines a dedicated authenticator in the mobile node's home network. The effectiveness and performance of this approach need to be analyzed.

Requirements. Information provided by the MIIS must contain not only the address of the target authenticator but also the address of the entity that is able to forward pre-authentication packets to the target authenticator, for example, the address of the home authenticator or the authenticator located in the third-party network.

Issue 1.8. In several scenarios pre-authentication with the target authenticator is not possible. This may happen when:

1. The target authenticator does not support direct pre-authentication for security reasons and the serving authenticator does not support forwarding of pre-authentication messages from the mobile node to the target authenticator (see Table C.1).
2. The target authenticator does not support pre-authentication.

Table C.1: Pre-authentication modes compatibility

TA	SA	HA*	Possibility of pre-authentication
Direct	All modes	All types	Yes
Indirect	Direct	Direct	No
Indirect	Indirect	Direct	Yes
Indirect	Direct	Indirect	Yes
Indirect	Indirect	Indirect	Yes

*HA may either the home or a third party proxy authenticator

C.2 Proposals for pre-authentication optimization in the inter-domain signalling

C.2.1 Pre-authentication with AAA server

This approach addresses Issue 1.1. To avoid the pre-authentication exchange with the mobile node and to decrease the traffic overhead the authentication server can respond with an "already authenticated" message if the pre-authentication session lifetime is not expired.

The pre-authentication result is cached on the AAA server and it may be pushed to an authenticator belonging to the same AAA domain. In other words, if there is a security association established between the AAA server and the authenticator, which is generally the case.

1. The mobile node is attached to the SA and it chooses a set of candidate authenticators.
2. The mobile node performs pre-authentication with all candidate authenticators.
3. The mobile node hands over to one of the candidate authenticators that becomes the new serving authenticator.
4. The mobile node sends a pre-authentication request to each new candidate authenticator.
5. If the mobile node is already authenticated with a candidate authentication server and its authentication session lifetime has not expired, the authentication server does not allow new pre-authentication, indicating that the session is valid. Whether the remaining lifetime should be sent to the mobile node and what is the minimum remaining session lifetime to reject new pre-authentication need to be studied.
6. The candidate authentication server pushes the MSK to the trusted authenticator that forwarded the previous pre-authentication request.
7. The mobile node hands over to the target authenticator that has already cached the MSK.
8. The mobile node and the target authenticator negotiate session keys.

Figure C.1 shows call flow for the proposed approach.

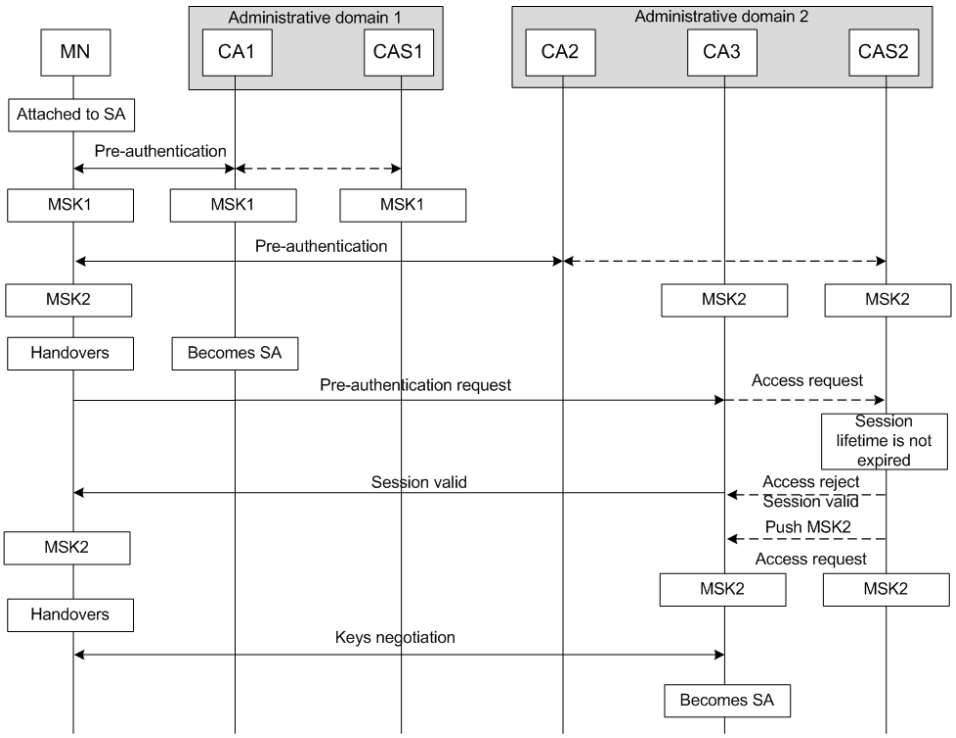


Figure C.1: Call flow for the pre-authentication with the CAS

Requirements:

1. Attribute-value pairs (AVPs) to inform the mobile node about unexpired session shall be defined [149].
2. The threshold value for the remaining session lifetime should be estimated.
3. The service provider network may be interested to make only one authenticator (or a limited set of authenticators) publicly accessible for security reasons.

Issues:

1. Pre-authentication latency with the target authenticator: $t_{auth} = RTT_{MN-TA}$
2. Pre-authentication latency with the target AAA server: $t_{auth} = RTT_{MN-TA} + RTT_{TA-AS}$
3. If the AAA server is not located in the access network (Remote AAA server) the RTT between the target authenticator and the authentication server is longer.

C.2.2 Using dedicated authenticator for pre-authentication

For security reasons an administrative domain may define pre-authentication support only for a small subset of authenticators (one at least). This dedicated target authenticator serves to ensure pre-authentication between the mobile node and the authentication server. When the MN handovers to the target authenticator belonging to the same AAA domain the MSK generated as a result of pre-authentication is pushed to the target

authenticator, the state of the mobile node changes to "authenticated", and the mobile node and the target authenticator negotiate session keys. Figure D.2 shows a call flow for the described scenario.

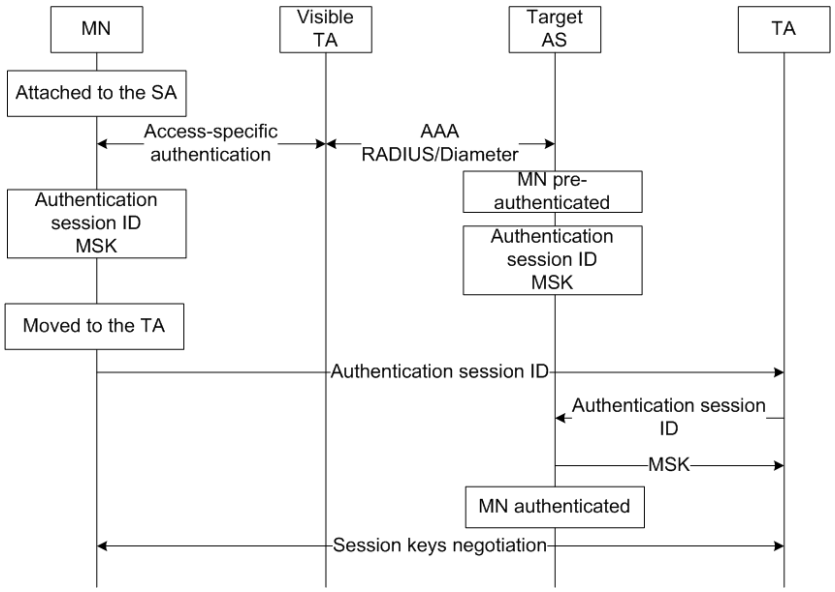


Figure C.2: Call flow for the pre-authentication with using dedicated TA

Requirements:

1. The entity performing the functionality of an authenticator shall be defined
2. Transport between the mobile node and this authenticator shall be defined.
3. A transport to be used between the authenticator and the mobile node should be defined.
4. It is desirable to push the MSK to the target authenticator before the mobile node handovers to the target authenticator (proactive mode).
5. The mechanism for mapping between addresses of the dedicated target authenticator and the target authenticator should be defined.
6. The authenticator that forwards pre-authentication requests must be either trusted for the target authentication server or it must be able to communicate with the authentication server that can proxy requests to the target authentication server.

C.3 MIH level security

C.3.1 General requirements

1. Both the MIH User and network entity may be the subject of an attack, therefore purposes of the security are:
2. MIH user protection from a fake MIH Information Service (MIH IS);

3. MIH IS protection form malicious users
4. Information received by the MIH User from the MIH Function is used to perform the next steps, and hence it is critical to protect it from alteration, modification, and provide message origin authentication. Due to short battery life at the mobile node it is essential to avoid processing of fake information by the mobile node.

The following issues need to be addressed and use cases need to be provided in regards to this topic:

1. Security of MIHF discovery
    a. There are two kinds of transport mechanisms: the first one is the lower layer transport (L2) and the second is the higher layer transport (L3).
    b. MIHF discovery: over media-specific L2 or L3 mechanism
    c. MIH Capability discovery – either over MIH or over media-specific broadcast messages
    d. To choose the set of candidate networks the mobile node must use only trusted and verified information.
    e. Mutual authentication of MIH peer nodes
    f. No authentication is assumed in the process of MIHF discovery and MIH Capability discovery is assumed in the current 802.21 draft.
    g. MIH pairing, from the mobile node's point of view, means authorization for the MIHF to send commands. Hence, the mobile node authorizes some important actions to an unauthenticated entity.
    h. MIHF registration assumes only identification of peers but it assumes any authentication and any means for integrity protection and message authentication of commands and events sent.
2. MIH access control
    a. The user should be able to select the most reliable IS among all available;
    b. After authentication different users are allowed to access different services.
    c. Per-user management of access rights is
        i. Costly;
        ii. Users may not be known in advance (if belonging to a different administrative domain);
        iii. User may not disclose its identity to the visiting network;
    d. Role-based management of access rights may be implemented instead.
    e. The role may be based on the user's state (unauthenticated/authenticated) or subscription (home/visiting).

C.3.2 Security of MIH Protocol

Regarding security of MIH Protocol the following requirements should be taken into consideration:

1. Re-using existing transport protocols
2. Re-using existing solutions for authentication, confidentiality, message authentication and integrity providing.

Due to the different nature of information provided by different mobility services (IS provides static information and Event Services and Command Services provide dynamic information about changes on the link) services-specific security requirements differ.

Information Service

1. Discovery may operate both within and outside administrative domain boundaries.
2. Definition of different sets of information available for users in authenticated and non-authenticated states;
3. Event Service and Command Service
4. Mutual authentication between the MIHF and the MIH User (simple authentication is not sufficient, particularly in the event of communication with the remote MIHF);
5. Secure channel establishment;
6. Providing confidentiality, integrity protection and message origin authentication.

The current 802.21 draft does not specify the location of the mobility services. In this way, the Information, Event and Command Services may be located in the serving, candidate or home network, or can even be managed by the third party authority. The [150] describes four scenarios for communication between the mobile node and Mobility Services (MoS).

Use Case 1. This Use Case corresponds to a scenario where the mobile node and the mobility service are located in the mobile node's home domain. In this situation the mobile node trusts the mobility service and it may use either L2 technology-specific mechanism of DHCP/DNS for mobility service discovery.

It is not always possible to establish link-layer security between two MIHF (for example, the mobile node and the mobility service are located in different networks). In this case L3 and upper layer security should be implemented.

Even if the mobile node and the mobility service are located in the same domain some issues need to be considered. One domain may include several networks spaced geographically. Information Service and, for example, Event Service may be located in different networks belonging to the same domain. The following assumptions are made:

1. Information, command and event have security associations established and

2. Data integrity and confidentiality is guaranteed for messages exchanged between mobility services belonging to the same administrative domain.

In this case only link-layer security between the mobile node and the current point of attachment should be established to let the mobile node and the mobility service communicate securely. Otherwise, L3 transport security should be established between the MN and each node providing mobility services.

Use case 2. This Use Case considers a scenario in which the mobile node and MIH Function are located in the visited domain. It is assumed that link-layer security is established between the mobile node and its point of attachment and L3 or upper level security is established between different network entities supporting the MIH Function. After performing mutual authentication between the mobile node and the visited network's AAA server the mobile node trusts network entities to send mobility related information and commands, and the network authorizes the mobile node to access mobility services according to the authorization profile associated with the mobile node.

Use Case 3. This Use Case describes a scenario where the mobile node is located in the visited network and the mobility service is located in the home network. If the mobile node is not pre-configured with the IP address of the MIH Function, it should be able to discover the MIHF in a secure manner. When the mobile node communicates with the Information Service, only server-to client authentication is required, while mutual authentication should be performed between the mobile node and Event Service / Command Service prior to MIH Function registration.

Use case 4. The mobile node is located in the visited or in the home network and the mobility service is located in the 3rd party network. L3 or upper level security should be implemented in this case. In order to establish trust between the mobile node and the MIH Function, the MIHF providing Information Service should authenticate itself to the mobile node. In such a way the mobile node makes decisions based on the information provided by the Information Service. To avoid communication between the MIH Function and an unauthorized mobile node, the mutual authentication should be performed between the mobile node and the MIH Function providing Event Service and Command Service.

C.3.3 Potential approaches

MIH Node discovery

The document [150] provides solutions for MIH Function discovery. If DHCP is used for node discovery, it is recommended to use the DHCP authentication option (RFC3118). This solution provides mechanisms both for node authentication and message authentication.

1. If DNS is used, it is recommended to use DNSSES (RFC 4033).
2. MIH transport security.
3. In a case when a reliable transport protocol such as TCP is used for connection between two MIHF peers, TLS (RFC 4366) should be used for data integrity and confidentiality.

4. In a case where unreliable transport protocol is used for connection between two MIH Function peers, DTLS (RFC 4347) may be used.
5. For generic IP level security, IPSec (RFC 2401) may be used if neither transport level security for a specific transport is available nor server-only authentication is required.

MIH –to-MIH authentication

Using FQDN and NAI as MIHF ID is applicable in all mentioned scenarios. IP address may be used as MIHF ID in a case where the mobile node and the mobility service are located in the same network (Use Case 1 and Use Case 2) or the IP address of mobility service is pre-configured on the mobile node.

Choosing the Information Service

The current 802.21 draft says: "The Media-Independent Information Service provides a framework and corresponding mechanisms by which a MIH Function entity may discover and obtain network information existing within a geographical area to facilitate the handovers. The obtained information may be used in conjunction with user and network operator policy."

Issue 2.1 is related to the choice of the Information Service. The current 802.21 draft does not specify the location of the Information Service. This way, the Information Service may be located in the serving, candidate or home network or the third party authority can even manage it. To choose the set of candidate networks the mobile node must use only trusted and verified information. Authorities providing Information Service may compete. This scenario causes issues when a mobile node uses information provided by Information Services in different candidate networks or the mobile node has multiples subscriptions.

Example: the mobile node is subscribed in two home networks. In the neighbourhood of its current network of attachment there can be partners of both home networks. To make the optimal choice, the mobile node should have access to all provided information (the home network may not provide information about prices of services in the partner network).

The mobile node may receive contradictory or conflicting information. That is why it is desirable to define some trust rating for the Information Service. This trust rating may be based on the previous experience: it is positive when the information provided was correct and it is negative if information provided was not correct. For handover decision making the mobile node chooses the set of Information Service with the highest rating.

Issue 2.2: The current 802.21 draft [30] specifies: "It is important to note that, with certain access networks a mobile node should be able to obtain IEEE 802.21 related information elements before the mobile node is authenticated with the point of attachment." In order to protect the user from receiving wrong information, the Information Service should be authenticated to the user (MIH Function-to-user authentication).

RELATED PUBLICATIONS

International conferences

1. M. Komarova and M. Riguidel, "Adjustable Trust-Based Access Control". *To appear in the Proceedings of The 5th International Conference on Autonomic and Trusted Computing (ATC-08). Oslo, Norway*

2. M. Komarova, M. Riguidel, Optimized Ticket Distribution Scheme for Fast Re-authentication Protocol (FAP). *ACM Q2SWinet Proceedings. Crete, Greece. 2007*

3. M. Komarova, M. Riguidel, A. Hecker, Fast re-Authentication Protocol for Inter-Domain Roaming *IEEE PIMRC'07 Proceedings. Athens, Greece, September 2007*

4. M. Komarova, M. Riguidel, Secure User's Mobility:State of the Art *4G & B3G DESIGN BOOK, GMC'2006, Beijing, China, October 2006*

5. M. Komarova, M. Riguidel, Wireless Network Architecture to Support Mobile Users. *ICETE 2006 WINSYS Proceedings. Portugal, August 2006. pp.325-330*

Journal article

6. M. Komarova and M. Riguidel, Secure User's Mobility: the current situation, *China Communications Journal. Special Issue on Wireless Communications, vol. 4, n° 1, pp. 95-104, February 2007*

Reports

7. M. Komarova, "Problem Statement for Authentication Signalling Optimization". *IEEE 802.21 MIHS Project; DCN 21-07-0387-00-0000.* November 2007

8. Subir Das, Marc Meylemans, Yoshihiro Ohba, Lily Chen, Nadia Golmie, Maryna Komarova, Michael Williams, Shubhranshu Sinha. "Technical Requirements document for MIH Security". *IEEE 802.21 MIHS Project, DCN 21-08-0012-00-0sec.* January 2008

BIBLIOGRAPHY

[1] IEEE Standard 802.11, “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications”, 1999 Editions, 1999.

[2] IEEE Standard 802.16, “Air Interface for Fixed Broadband Wireless Access Systems”, 2004 Edition, 2004

[3] IEEE Standard 802.16e, „Air Interface for Fixed Broadband Wireless Access Systems. Amendment 2: Physical and Medium Access Control layers for Combined Fixed and Mobile Operations in Licensed Bands and Corrigendum 1“. 2006 Edition.

[4] GSM 01.02 (ETR 99): "European digital cellular telecommunications system (Phase 2); General Description of a GSM PLMN". 1999

[5] ETSI GSM 04.60 “Digital Cellular Telecommunications System: General Packet Radio Service (GPRS): (phase 2+): Overall description of the GPRS Radio Interface (Um)”, ver 0.9.0, 18 September 1996.

[6] ETSI TS 123 101 V7.0.0. “Universal Mobile Telecommunications System (UMTS);General UMTS Architecture (3GPP TS 23.101 version 7.0.0 Release 7)” 2007

[7] International Telecommunication Union website: http://www.itu.int

[8] European Telecommunications Standards Institute specifications : http://www.etsi.org

[9] IMS – IP Multimedia Subsystem, Ericsson White Paper, 2004 (Retrieved from http://www.ericsson.com/technology/whitepapers/ims_ip_multimedia_subsystem.pdf)

[10] “Methods for subjective determination of transmission quality”. ITU Recommendation P.800. 1996

[11] Perceptual Evaluation of Speech Quality (PESQ). ITU-T recommendation P.862, May 2000

[12] The E-model, a computational model for use in transmission planning.. ITU-T Recommendation G.107, March 2003

[13] Héctor Velayos, Gunnar Karlsson. “Techniques to Reduce IEEE 802.11b MAC Layer Handover Time”. In: IEEE International Conference on Communication (ICC), June 2003.

[14] IEEE Standard 802.11i “Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. Amendment 6: Medium Access Control (MAC) Security Enhancements”. 2004 Edition. 2004

[15] S. Thomson, T. Narten. “IPv6 Stateless Address Autoconfiguration”. Request for Comments 2462. December 1998

[16] R. Droms. “Dynamic Host Configuration Protocol”. Request for Comments 2131. March 1997

[17] C. Perkins “IP Mobility Support”, Request for Comments 2002, October 1996

[18] J. Risenberg, H. Schulzrinne. “SIP: Session Initiation Protocol”. Request for Comments 3261. June 2002

[19] Ivan Martinovic, Frank A. Zdarsky, Adam Bachorek, and Jens B. Schmitt: “Measurement and Analysis of Handover Latencies in IEEE 802.11i Secured Networks”. In Proceedings of the 13th European Wireless Conference (EW2007), Paris, France. April 2007.

[20] Arunesh Mishra, Minho Shin, William Arbaugh. “An Empirical Analysis of the IEEE 802.11 MAC Layer Handoff Process”. In: ACM Computer Communication Review, Apr. 2003, 33(2): 93~102

[21] Jon-Olov Vatn. “An experimental study of IEEE 802.11b handover performance and its effect on voice traffc”, Technical Report TRITA-IMIT-TSLAB R 03:01, Telecommunication Systems Laboratory, Department of Microelectronics and Information Technology, KTH, Royal Institute of Technology, Stockholm, Sweden2003

[22] IETF “Requirements for Internet hosts”. Request for Comments 1122. October 1989

[23] ITU-T, "General Characteristics of International Telephone Connections and International Telephone Circuits: One-Way Transmission Time", ITU-T Recommendation G.114 1998.

[24] J. Manner, M. Kojo. “Mobility Related Terminology”. Request for Comments: 3753, June 2004

[25] S. Hares, D. Katz. “Administrative Domains and Routing Domains: A model for routing in the Internet”. Request for Comments 1136 1989

[26] IEEE P802.11r “Fast Roaming/ Fast BSS Transition”. Draft D0.9.January 2008

[27] IEEE Trial-Use Recommendation Practice for Multi-Vendor Access Point 802.11f “Interoperability via an Inter-Access Point Protocol Across Distribution System Supporting IEEE 802.11 Operation”. 2003 Edition

[28] IEEE P802.11u “Interworking with External Networks”. Draft D2.0. January 2008

[29] IEEE P802.11e “MAC Enhancement for Quality of Service”. Draft D13.0. July 2005

[30] IEEE P802.21 “Media Independent Handover Services”. Draft D7.1. August 2007

[31] C. Perkins, “IP Mobility Support for IPv4”, Request for Comments 3344, August 2002

[32] D. Johnson, C. Perkins, J. Arkko , “Mobility Support in IPv6”, Request for Comments 3775, June 2004

[33] C.Blondia at al, “Performance evaluation of Layer 3 Low Latency Handoff Mechanisms”, Mobile Networks and Applications 9, 2004

[34] R. Koodli, “Fast Handovers for Mobile IPv6”, Request for Comments: 4068. July 2005

[35] H. Soliman et al., “Hierarchical Mobile IPv6 Mobility Management (HMIPv6),” Internet draft, IETF, draft-ietfmipshop-hmipv6-02.txt, June 2004, work in progress.

[36] S.Das et al., “IDMP: An Intra-Domain Mobility Management Protocol for Next Generation Wireless Networks”, IEEE Wireless Magazine, October 2002

[37] A. Campbell et al., “Cellular IP,” draft-ietf-mobileip-cellularip-00.txt, IETF, January 2000,Work in Progress

[38] R. Ramjee et al., “IP micro-mobility support using HAWAII,” draft-ietf-mobileip-hawaii-01.txt, July 2000, Work in Progress

[39] R. Moskowitz, “Host Identity Protocol Architecture”, draft-ietf-hip-arch-03, August 1, 2005, work in progress.

[40] ANSI T1.244-1995 -- Operations, Administration, Maintenance, and Provisioning (OAM&P)-Interface Standards for Personal Communications Services

[41] The NIST Handbook, Special Publication 800-12, An Introduction to Computer Security.

[42] Compact Oxford English Dictionary of Current English. ISBN-13: 978-0-19-861022-9. 1264 pages, 2005.

[43] Gambetta, Diego (2000) ‘Can We Trust Trust?’, in Gambetta, Diego (ed.) Trust:Making and Breaking Cooperative Relations, electronic edition, Department of Sociology, University of Oxford, chapter 13, pp. 213-237

[44] S. Marsh, ”Trust and reliance in Multi-agent systems: a preliminary report”, MAAMAW’92, Italy, 1992

[45] Yahalom, R.; Klein, B.; Beth, T. “Trust relationships in secure systems-a distributed authentication perspective“,Research in Security and Privacy, 1993. Proceedings., 1993 IEEE Computer Society Symposium on 24-26 May 1993 Page(s):150 - 164

[46] T. Beth, M. Borcherding and B. Klein, “Valuation of Trust in Open Networks”, In: Proceedings of European Symposium on Research in Computer Security (ESORICS), pp. 3-18, 1994, Springler-Verlag

[47] Lea Viljanen, “Towards an Ontology of Trust”. In: Proceedings of Second International Conference, TrustBus 2005, Copenhagen, Denmark, August. Pp. 185-174 2005

[48] Nikita Borisov, Ian Goldberg and David Wagner. “Intercepting Mobile Communications: The Insecurity of 802.11.” 7th Annual International Conference on Mobile Computing and Networking, ACM Mobicon 2001

[49] William A. Arbaugh, “An Inductive Chosen Plaintext Attack against WEP/WEP2”, doc IEEE802.11-01/230, 2001

[50] CNN.com, “'Off-the-shelf' hack breaks wireless encryption”, August 11, 2001

[51] IEEE, Standards for local and metropolitan area networks: Standard for port based network access control, IEEE Standard P802.1X, October 2001

[52] B. Aboba et al., “Extensible Authentication Protocol (EAP)”, Request for Comments 3748, June 2004

[53] Rigney, C., Willens, S., Rubens, A., Simpson, W. (June 2000). Remote Authentication Dial In User Service (RADIUS). RFC 2865. Retrieved from www.ietf.org.

[54] R. Housley, W. Ford, W. Polk, D. Solo, “Internet X.509 Public Key Infrastructure Certificate and CRL Profile”. Request for Comments 2459. January 1999

[55] ETSI. TS 33.102 V 7.1.0: Universal Mobile Telecommunication system (UMTS); 3G Security; Security Architecture (3GPP TS 33.102 Version 7.1.0 Release 7), from http://www.3gpp.org/ftp/Specs/html-info/33102.htm. 2006

[56] ETSI GSM 03.20: "Digital cellular telecommunications system (Phase 2+); Security related network functions". February 1992

[57] 3GPP TS 33 105 V 7.0.0: Universal Mobile Telecommunication system (UMTS); Cryptographic Algorithms Requirements (3GPP TS 33.105 version 7.0.0 Release 7). 2007

[58] J. Franks et al, “HTTP Authentication: Basic and Digest Access Authentication”. Request for Comments: 2617. June 1999.

[59] Dan Forsberg, Yoshihiro Ohba, Basavaraj Patil, Hannes Tschofenig, Alper Yegin, “Protocol for Carrying Authentication for Network Access (PANA)”. draft-ietf-pana-pana-18.txt. 2005

[60] E. Damiani, S. D. C. di Vimercati, and P. Samarati, "Managing multiple and dependable identities," IEEE Internet Computing, vol. 7, no. 6, pp. 29--37, November/December 2003.

[61] A. Jøsang, J. Fabre, B. Hay, J. Dalziel and S. Pope. Trust Requirements in Identity Management. Proceedings of the Australasian Information Security Workshop (AISW'05), Newcastle, Australia, January-February 2005

[62] Abhilasha Bhargav-Spantzel, Anna C. Squicciarini and Elisa Bertino. “Trust Negotiation in Identity Management”. In: IEEE Security & Privacy, Volume 5, Issue 2, pp. 55-63. March-April 2007

[63] Mo Li, Kumbesan Sandrasegaran, Xiaoan Huang: Identity Management in Vertical Handovers for UMTS-WLAN Networks. In: IEEE ICMB 2005: 479-484

[64] Aboba, B., Blunk, L., Vollbrecht, J., Carlson, J., and H. Levkowetz, "Extensible Authentication Protocol (EAP)", RFC 3748, June 2004.

[65] T. Berners-Lee, R. Fielding , L. Masinter. “Uniform Resource Identifier (URI): Generic Syntax”. IETF Request for Comments: 3986. January 2005

[66] D.Halasz, “IEEE 802.11i and wireless security”, august 2004// www.Embedded.com

[67] “Secure Computer Systems,” ESD-TR-73-278, Mitre Corporation; v I and II (Nov 1973), v III (Apr 1974).

[68] K Biba, Integrity Considerations for Secure Computer Systems, Mitre Corporation MTR-3153 (1975).

[69] Ross J. Anderson. “Security Engineering: A Guide to Building Dependable Distributed Systems”. Willey Edition, ISBN: 978-0-471-38922-4, 640 p., 2001

[70] “Identity based control”, ConSentry networks White paper, 2006

[71] Sandhu, R.S.; Coyne, E.J.; Feinstein, H. L.; Youman, C.E.;” Role-based access control models”. In IEEE Computer, Volume 22, Issue 9: 38-47, Feb.1996

[72] David F.Ferraiolo, Ravi Sandhu, Serban Gavrila, D. Richard Kuhn and Ramaswamy Chandramouli. “Proposed NIST Standard for Role-Based Access Control”. In: ACM Transactions on Information and System Security, Vol. 4, No. 3, August 2001, Pages 224–274.

[73] Ninghui Li, Mahesh V. Tripunitara. “Security analysis in role-based access control”. In: ACM Transactions on Information and System Security, Vol. 9, No. 4, November 2006, Pages 391–420.

[74] Gail-Joon Ahn, Hongxin Hu. “Towards realizing a formal RBAC model in real systems”. In: Proceedings of the 12th ACM symposium on Access control models and technologies (SACMAT’07), June 2007

[75] A. Josang, “The Beta Reputation system”, Proceedings of the 15th Bled Conference on Electronic Commerce, Bled, Slovenia, 17-19 June 2002.

[76] M. Srivasta, Li Xiong, Ling Liu, “Trust Guard: Countering Vulnerabilities in Reputation Management for decentralized Overlay Networks”, WWW 2005

[77] Nathan Griffits, “Task Delegation using Experience-Based Multi-Dimensional trust”, AAMAS’05.

[78] S. Park et al, “Resilient Trust Management for Web Service Integration”. IEEE International Conference on Web Services, 2005. ICWS 2005.

[79] SECURE project’s website: http://www.dsg.cs.tcd.ie/dynamic/?category_id=-30

[80] N.Dimmock et al, “Risk models for Trust-based Access control (TBAC)”. In Proc. 3rd Annual Conference on Trust Management (iTrust), pages 364–371, May 2005. 2004

[81] N.Dimmock et al, “Using Trust and Risk in Role-Based Access Control Polilcies”. In Proc. Symposium on Access Control Models and Technologies (SACMAT), pages 156–162, June 2004.

[82] Huu Tran, Michael Hitchens, Vijay Varadharajan, Paul Watters, “A Trust based Access Control Framework for P2P File-Sharing Systems”. In: 38th Hawaii International Conference on System Sciences (HICSS-38 2005). January 2005

[83] Naouel Ben Salem et al, “Reputation-based Wi-Fi Deployment Protocols and Security Analysis”, WMASH’04, 2004

[84] Trung Dong Huynh et al, “An Integrated Trust and Reputation model for Open Multi-Agent Systems”. Journal of Autonomous Agents and Multi-Agent Systems, 13 (2). pp. 119-154. 2004

[85] K. Krukow, M. Nielsen, V. Sassone, “A Framework for Concrete reputation-systems with Applications to History-Based Access Control”. CCS’05.

[86] Y. Ohba, “EAP pre-Authentication Problem Satement”. draft-ietf-hokey-preauth-ps-02. Work in progress February 2008

[87] IEEE 802.21 Technical Requirements document for MIH Security (21-08-0012-00-0sec_MIH_Security_Technical_Report). January 2008

[88] MIPSHOP WG “mobility Services Transport protocol Design”. draft-melia-mipshop-mstp-solution-01. Work in progress. November 2007.

[89] Mattheß, M., C.O. Krauß, K.M. Bayarou, C. Eckert, A.R. Prasad and P. Schoo: Identification of Security Requirements in WLAN-WLAN Inter-Domain Handovers. In: The 8th International Symposium on Wireless Personal Multimedia Communications WPMC 2005, Aalborg, Denmark, September 17-22 2005

[90] B. Aboba, M. Beadles. “The Network Access Identifier”. IETF Request for Comments: 2486. January 1999

[91] 3GPP Technical Specification 35.202: Design of the KASUMI Block Cipher.

[92]] R. Housley, W. Ford, W. Polk and D. Solo, “Internet X.509 Public Key Infrastructure Certificate and CRL Profile”, Request for Comments 2459, April 2002

[93] OpenID project website http://openid.net/

[94] Liberty alliance website: http://www.projectliberty.org/

[95] Shibboleth project website: http://shibboleth.internet2.edu/

[96] L. Reznik, “Which models should be applied to measure computer security and information assurance?”. In: The IEEE International conference of Fuzzy Systems, 2003

[97] D. Stanley, J. Walker, B. Aboba. “Extensible Authentication Protocol (EAP) Method Requirements for Wireless LANs”. IETF Request for comments 4017. 2005

[98] H. Koshutanski, F. Martinelli, P. Mori and A. Vaccarelli, “Fine-grained and History-based Access Control with Trust Management for Autonomic Grid Services”. 2006

[99] Michael Chuang, Suronapee Phoomvuthisam, James B. D. Joshi “An Integrated Framework for Trust-Based Access Control for Open Systems”. International Conference on Collaborative Computing: Networking, Applications and Worksharing, pages 1-12. 2006

[100] Guo Ya-Jun, Hong Fan, Zhang Qing-Guo, Li Rong. “An Access Control Model for Ubiquitous Computing Application”, Mobile Technology, Applications and Systems, 2005 2nd International Conference, 15-17 Nov. 2005 Page(s):1 - 6

[101] Tchepnda, C., Riguidel, M. “Distributed Trust Infrastructure and Trust-Security Articulation: Application to Heterogeneous networks”. In: Proceedings of Advanced Information networking and Applications. AINA 2006

[102] T. Beth, M. Borcherding, B. Klein, “Valuation of Trust in Open Networks”. In D. Gollmann, editor, ESORICS 94, Brighton, UK, November 1994.

[103] Ajay Ravichandran, Jongpil Yoon, “Trust management with delegation in grouped peer-to-peer communities”. In: Proceedings of the eleventh ACM symposium on Access control models and technologies, Pages: 71 – 80. 2006

[104] A. Hecker, H.Labiod, “Pre-authentication signalling in Wireless LANs using 802.1X access control”. IEEE GLOBECOM 2004, Dallas, TX, USA.

[105] Mobile and Wireless Systems beyond 3G. Project “Ambient Networks Phase 2“. D7-A.2 Draft System Description. FP6-CALL4-027662-AN P2/D07-A2. 2007.

[106] Unlicensed Mobile Access (UMA); Architecture (Stage 2). Technical Specification R1.0.1.2004

[107] J. Arkko, H. Haverinen. “Extensible Authentication Protocol Method for 3rd Generation Authentication and Key Agreement (EAP-AKA)”. Informational Request for Comments 4187. January 2006

[108] H. Haverinen, J. Salowey. “Extensible Authentication Protocol Method for Global System for Mobile Communications (GSM) Subscriber Identity Modules (EAP-SIM)”. Informational Request for Comments 4186. January 2006

[109] S.Pack and Y. Choi, “Fast Inter-AP handover using predictive authentication scheme in a Public wireless LAN”. In/ Proc. Of Networks 2002, Aug. 2002

[110] S.Pack et al, “A Selective Neighbour Caching Scheme for fast handoff in IEEE 802.11 Wireless Networks”, ACM SIGMOBILE Mobile Computing and Communications Review, 2005

[111] A.Mishra et al, “Proactive key Distribution Using Neighbour Graphs”, IEEE Wireless communications, February 2004

[112] Kassab, M., Belghith, A., Bonnin, J.-M., Sassi, S. (2005, October, 13). Fast Pre-Authentication Based on Proactive Key Distribution for 802.11 Infrastructure Networks. WMuNeP’05. The ACM Digital Library.

[113] Kassab, M., Bonnin, J.-M, Belghith, A. “Fast and Secure Handover in WLANs: An evaluation of the signalling overhead”. In: IEEE CCNC 2008 proceedings.

[114] M.S.Bargh at al, “Fast authentication Methods for handovers between IEEE 802.11 Wireless LANs”, WMASH’04, October 1, 2004

[115] J. Loughney, C. Perkins and R. Koodli. “Context transfer Protocol (CXTP)”. Experimental Request for Comments 4067. July 2005

[116] N. Aboudagga, M. Eltoweissy and J.-J Quisquater, “Fast Roaming Authentication in Wireless LANs”, 1st Benilux Workshop on Information and System Security, Antwerpen, Belgium, November 2006

[117] IETF HOKEY Working Group webpage http://www.ltsnet.net/ietf/hokey/

[118] M. Nakhjiri. “Use of EAP-AKA, IETF HOKEY and AAA mechanisms to provide access and handover security and 3G-802.16m interworking”. In: Proceedings of PIMRS’07. 2007

[119] IETF Seamoby working group website: http://merlot.tools.ietf.org/wg/seamoby/

[120] B. Gaabab, D. Binet and J.-M. Bonnin. “Authentication Optimization for Seamless Handovers”. 10th IFIP/IEEE International Symposium on Integrated Network Management, 2007.

[121] Yegin, A., Ed., Ohba, Y., Penno, R., Tsirtsis, G., and C. Wang, "Protocol for Carrying Authentication for Network Access (PANA) Requirements and Terminology", Work in Progress, August 2004.

[122] D. Forsberg, J. Bournelle, R. Marin Lopez. “PANA Mobility Optimizations with Session Keys Context (SKC)”. draft-forsberg-pana-skc-00. 2005

[123] J. Kohl, C. Neuman, “The Kerberos Network Authentication Service V.5”. Request for Comments 1510. September 1993

[124] Z. Hong , H. Rui, Y. Man, K. Zhigang, Q. Hualin, “A Novel Fast Authentication Method for Mobile Network Access”, International Conference for Young Computer Scientists (ICYCS), Harbin, October, 2003

[125] H. Wang, A. R. Prasad, “Fast Authentication for Inter-domain Handover”, in Proc. of International Conference on Telecommunications (ICT’04), Fortaleza, Brazil, August 1-7, 2004.

[126] M. Long, Ch.-H. “John” Wu, J. D. Irwin, “Localized Authentication for Wireless LAN Inter-network Roaming”, Communications, IEEE Proceedings-Volume 151, Issue 5, 24 Oct. 2004 Page(s):496 – 500

[127] S. G. Polito, H.Schulzrinne, “Authentication and Authorization Method in Multi-domain, Multi-provider Networks”.In/ Proceedings of Next Generation Internet Networks, 3rd EuroNGI Conference. 2007

[128] Y. Ohba, S. Das and A. Dutta, “Kerberized Handover Keying: A media-independent handover key management architecture”. In: Proceedings of MobiArch’07, August 2007, Japan

[129] Pho Duc Giang, Le Xuan Hung, Sungyoung Lee, Young-Koo Lee and Heejo Lee. “A Flexible Trust-Based Access Control Mechanism for Security and Privacy Enhancement in Ubiquitous Systems”. IEEE MUE'07, April 2007

[130] Sudip Chakraborty, Indrajit Ray, “TrustBAC - Integrating Trust Relationships into the RBAC Model for Access Control in Open Systems”, Proceedings of the eleventh ACM symposium on Access control models and technologies, 2006, Pages: 49 - 58

[131] Y.Matsunaga et al, “Secure Authentication System for Public WLAN Roaming”, WMASH’03, California, USA, - September 2003

[132] Matsunaga, Y., Merino, A.S., Suzuki, T., Katz, R.H. (September 2003). Secure Authentication System for Public WLAN Roaming. WMASH’03. Retriewed from http:\\berkeley.edu/paper

[133] Das, S., Patil, B., Soliman, H., Yegin, A. (2003, April, 28). Problem Statement and Usage Scenarios for PANA. draft-ietf-pana-usage-scenarios-06.txt. Retrieved from www.ietf.org

[134] Bargh, M.S., Hulsebosch, R.J., Eertink, E.H., Prasad, A., Wang, H., Schoo, P. (2004, October, 1). Fast authentication Methods for handovers between IEEE 802.11 Wireless LANs. WMASH’04. The ACM Digital Library.

[135] Patil, B., Tschofenig, H., Yegin, A. (2005, October, 21) PANA mobility optimizations. draft-ietf-pana-mobopts-01. Retrieved from www.ietf.org.

[136] Parthasarathy, M. (March 2005). Protocol for Carrying Authentication and Network access (PANA) Threats Analysis and Security requirements. RFC 4016. Retrieved from www.ietf.org

[137] Loughney, J., Nakhjiri, Ed.M., Perkins, C., Koodli, R. (July 2005). Context Transfer Protocol (CXTP). RFC 4067. Retrieved from www.ietf.org

[138] Aboba, B., Beadles, M.(January 1999). The Network Access Identifier. RFC 2486. Retrieved from www.ietf.org

[139] D. Whiting, R. Housley and N. Ferguson. “Counter with CBC-MAC (CCM)”. Request for Comments 3610. September 2003

[140] Secure Hash Signature Standard (SHS) (FIPS PUB 180-2), Federal Information Processing Standards Publication 180-2, August 2002

[141] D. Stanley, J. Walker and B. Aboba, “Extensible Authentication Protocol (EAP) Method Requirements for Wireless LANs”, Request for Comments 4017, March 2005

[142] Krawczyk, H., Bellare, M., and R. Canetti, "HMAC: Keyed-Hashing for Message Authentication", RFC 2104, February 1997.

[143] FreeRadius.org

[144] open1x.sourceforge.net

[145] www.wireshark.org

[146] Luis Cabaral, Ali Hortaçsu. “The dynamics of seller reputation: evidence from eBay”. In: eScholarship Repository, University of California. http://repositories.cdlib.org/berkeley econ221/fall2005/4. 2005

[147] www.omnetpp.org

[148] 21-07-0122-04-0000-Security_proposal.ppt, “Security Optimization During Handovers: 802.21 SG Proposal”.

[149] B. Aboba, “RADIUS Attributes for WLAN” . IETF draft draft-aboba-radext-wlan-00, work in progress, July 2005

[150] Mipshop WG “Mobility Services Transport Protocol Design”. IETF draft draft-melia-mipshop-mstp-solution-01, work in progress. November 2007