UNIVERSITY OF WARMIA AND MAZURY IN OLSZTYN

Technical Sciences

15(1)

PUBLISHER UWM

OLSZTYN 2012

The Technical Sciences is indexed and abstracted in BazTech (http://baztech.icm.edu.pl)

The Journal is also available in electronic form on the web site
http://wydawnictwo.uwm.edu.pl (subpage *Czytelnia*)

The print edition is the primary version of the Journal

PL ISSN 1505-4675

Address
ul. Jana Heweliusza 14
10-718 Olsztyn-Kortowo, Poland
tel.: +48 89 523 36 61
fax: +48 89 523 34 38
e-mail: wydawca@uwm.edu.pl

Ark. wyd. 15,4, ark. druk. 12,5, nakład 120 egz.
Druk – Zakład Poligraficzny UWM w Olsztynie
zam. nr 396

Contents

Environmental Engineering

Geodesy and Cartography

Information Technology

Spis treści

Inżynieria środowiska

Geodezja i kartografia

Technologie informacyjne

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

CULTIVATION OF *SALIX VIMINALIS* WILLOW AND POSSIBILITIES OF IMPROVING THE ENERGY BALANCE OF EASTERN POLAND VOIVODSHIPS

Adam J. Lipiński, Alicja A. Żejmo

Chair of Working Machines and Separation Processes
University of Warmia and Mazury in Olsztyn

Key words: willow biomass, *Salix viminalis* willow, eastern Poland, crop potential.

Abstract

The aim of this study was to analyze the possibility of improving eastern Poland energy balance by transforming the agricultural areas, destined for food production purposes, to crops with industrial purpose. The problem was analyzed in terms of perspectives, problems and challenges facing the eastern Poland voivodships in order to diversify energy sources. The direction was indicated, which, according to the performed analysis, gives the best chance of improving the present situation, in terms of energetic security and also improving the socio-economic conditions of the inhabitants of these regions. The expediency of exploring opportunities of producing energy from biomass, with particular reference to the *Salix viminalis* willow, was pointed out.

UPRAWA WIERZBY *SALIX VIMINALIS* A MOŻLIWOŚCI POPRAWY BILANSU ENERGETYCZNEGO WOJEWÓDZTW POLSKI WSCHODNIEJ

Adam J. Lipiński, Alicja A. Żejmo

Katedra Maszyn Roboczych i Procesów Separacji
Uniwersytet Warmińsko-Mazurski w Olsztynie

Słowa kluczowe: biomasa z wierzby, wierzba *Salix viminalis*, Polska Wschodnia, potencjał upraw.

Abstrakt

Celem pracy była analiza możliwości poprawy bilansu energetycznego Polski Wschodniej przez przekształcanie upraw rolniczych o przeznaczeniu żywnościowym na uprawy z przeznaczeniem przemysłowym. Zagadnienie przeanalizowano pod kątem perspektyw, problemów i wyzwań stojących przed województwami Polski Wschodniej w kwestii dywersyfikacji źródeł energii. Wskazano kierunek działań, które według przeprowadzonej analizy dają największe szanse poprawy obecnej sytuacji, zarówno w zakresie bezpieczeństwa energetycznego, jak i poprawy warunków socjalno-ekonomicznych mieszkańców omawianego regionu. Wskazano na celowość poszukiwań możliwości produkcji energii z biomasy, ze szczególnym uwzględnieniem wierzby *Salix viminalis*.

Introduction

Constantly rising prices of conventional energy sources (oil, natural gas, coal) prompt to search for alternative sources of energy. Renewable energy sources are becoming a substitute for its conventional sources (LESZ 2006). Poland, after joining the European Union (EU), was obliged to implement the Directive on the share of Renewable Energy Sources (RES) in the energy sector. Polish parliament approved the "Strategy for Development of Renewable Energy", which obliges to achieve 7% (in 2010) and 14% (in 2020) of the RES share in the country energy balance (NIEDZIÓŁKA, ZUCHNIARZ 2006, BIENIEK, ŻOŁNIERZ-RUSINEK 2008).

Currently, approximately 52% of energy used in Europe is imported from other continents and this share is prognosed to increase (DENISIUK 2006, *Charakterystyka obszarów wiejskich* 2008). According to DENISIUK (2006), as early as in 2015, 72% of energy will be imported to the EU. Given this, the idea of allocating agricultural lands for purposes of non-food production seems to be reasonable, especially since such actions may help to solve the problem of agricultural overproduction in eastern Poland, typically regarded as agricultural land (KUKUŁA, KRASOWICZ 2006, *Charakterystyka obszarów wiejskich*... 2008). Another appealing premise for the introduction of such crops is protecting the local population against the progressive impoverishment and exclusion, being effects of hidden unemployment (MYSZCZYSZYN 2001). In case of larger farms, it will allow diversifing their income and exploit uncultivated land. Non-agricultural activities of agricultural holdings are positive phenomenon. It reflects the process of adaptation of agriculture to market conditions and allows for the relative security in terms of possible crisis on markets for specific products (GĄSIOREK 2005, LESZ 2006).

Poland in the European Union is perceived as country with great possibilities of biomass production. It is estimated that up to 1.6 million ha of agricultural land can be allocated under that production. The current area of growing plants for solid fuels is estimated at about 10 thousand ha (VAN DAM et al. 2005, KUŚ et al. 2008). The use of biomass energy potential is dependent on many issues, among which the most important is economic factor, conditioning the formation of biomass projects. EU legislation regulates these issues in some detail (EU directives 2001, 2009). This applies to the principles of electricity market operation and methods of electric energy production. The percentage of energy from renewable sources in total energy production in Poland is 7.9% and is lower than the EU average, which is 10.3% (Eurostat 2008). It should be noted that as many as 94% of renewable energy produced in Poland is obtained from biomass (LESZ 2006, PAWLAK 2007). The degree of biomass use on a local scale depends largely on the correct identification of local production opportunities (SIEJKA et al. 2008).

Aim and scope of work

The aim of this study was to determine the natural and socio-economic determinants of agricultural development on eastern Poland territory in terms of non-agricultural farm activities and to search for opportunities to improve the energy balance of the eastern Poland voivodships.

The scope of work included the review of the challenges and opportunities facing the eastern Poland voivodships under the circumstances of future inevitable changes of the agrarian structure of Poland and in order to introducing alternative energy sources.

Problem justification

The development of technologies, as well as trends in research and development of energy-generating systems, and the priorities of the Commission of the European Union under the following acts: Green Paper (2000), the Kyoto Protocol (2005) and the European Directives (2001/77/WE, 2001/77/EC), tend to consider wider exploitation of renewable energy sources (ŚCIBISZ 2006, PIECHOCKI, BIERANOWSKI 2007). Many studies show a high probability of a future energy crisis, the differences concern only its occurrence time forecast (15 to 25 years) (ROSZKOWSKI 2008). Under the Polish energetic law, local authorities are required to develop plans for heat and energy supply. These plans should also include the use of locally available renewable energy sources.

The most common method of obtaining electrical energy is to use the power grid, covering almost the entire country (WALSKI 2007). For thermal energy, electric heaters or charcoal/gas stoves are used and, more and more frequently, biomass stoves (ŚCIBISZ 2006). Conventional energy sources may be replaced by its renewable sources and, in terms of integration with the European Union, active participation in the campaign of implementation of renewable energy can be an asset of Poland. The most important reason for promoting the production of biomass for energy purposes is the desire to tackle climate changes (SOŁOWIEJ 2006, ŚCIBISZ 2006, SZWEDZIAK 2010).

Changes in the era of economic development are forcing manufacturers and distributors of energy to explore new forms of fuels structure, especialy those more environmentally friendly. The use of biomass energy potential in both, the global and the local area, is dependent on many factors. The most important of these is the economic factor, conditioning the formation of biomass projects. The degree of using biomass on a local scale (e.g. a specified municipality) depends above all on the correct identification of possibility of local biomass production (GRZYBEK 2007, SIEJKA et al. 2008, SZWEDZIAK 2010). The shift of the individual farm from classical methods of meeting its energy needs (power grid, coal stoves) to the use of renewable energy requires the economic analysis of the entire project. It must be based on determining the energy needs of the specified farm purposed to modernization and the evaluation of energy resources available within it, but the specificity of individual farms, in terms of the character of production, as well as their location, particularly predispose them to use renewable energy sources (SOŁOWIEJ 2006, ŚCIBISZ 2006). The use of agricultural biomass in most cases is contradictory to the basic purposes of agriculture, implied as the production of food. To convince farmers to cultivating willow, it must be profitable, at least as profitable as cultivating cereals.

Energetic use of rural areas implies a threat of disturbances on food markets. Despite this, establishments of willow plantations are becoming increasingly popular topic (*Charakterystyka obszarów wiejskich*... 2008). Growing energetic crops is part of a complex system, which is the national economy. At the macro level, it should therefore be considered in many aspects, including the principles and stages of implementation of the strategy of obtaining energy from renewable sources (ROSZKOWSKI 2008). Unfortunately, decisions concerning energetic crops production are currently being made using the results of incomplete energetic and environmental analyzes, often under the influence of constantly changing economic regulation system (donations system).

Energy efficiency of wilow plantations depends mostly on the agrotechnical procedures performed there (LESZ 2006, GRZYBEK 2007, PAWLAK 2007, ROLA et al. 2007). At plantations surveyed by Kwaśniewski (2008), willow cultivation required high initial material and resource costs, which may pose an additional problem in the impoverished farms.

Much of the energy potential of biomass is lost in the overall balance of energy flow through the area. According to the study of the Board of the Warmia and Mazury voivodship (2005), the condition of the profitability of willow transport from the field to the point of combustion is fulfilled if the itinerary does not exceed 80 km, thus willow-growing farms should be located possibly close to the location of existing combustion installations. It is therefore appropriate to define the conditions for decentralized heat and electricity production from local energetic resources (SIEJKA et al. 2008, SZWEDZIAK 2010).

Deepening economic integration of Europe is advantageous for countries and regions more developed, but also create a series of threats to the the less developing and peripheral regions. Polish accession to the European Union has highlighted the problem of the differences in development between the countries lying on one continent, as well as interregional differences at the scale of the state, as well as at a whole Community scale. Depopulation trends are noted: population growth of area considered in whole is negative and slow exodus to other voivodships is observable (GUS 2004–2009, PAWLAK 2009). System transformation and industrial restructurization has resulted in declining the agricultural production potential (GUS 1995–2005). Country aid is directed to industrialized areas, to eliminate the negative effects of a changing economy. This resulted in a deepening interregional diversity (GREWIŃSKI 2008). Eastern Poland voivodships are characterized by the lowest level of per capita investment in the national economy in general (MIKOŁAJCZYK 2008). These conditions are reflected in the level of development of the eastern provinces, when compared to the rest of the country. It concerns the level of infrastructure development, as well as the attitude of their inhabitants to the enterprise, resulting in low levels of vocational activity (employment rate was 54.8% – GUS), low standard of living and low dynamics of economic development (www.mrr.gov.pl 2009). Introduction of energetic plantations in these areas can create employment opportunity in areas of construction, operation and service of renewable energy systems.

Characteristics of biomass from *Salix viminalis* willow

Energetic plants have a high annual growth, high calorific value, high resistance to pests or diseases and the relatively small soil requirements (NIEDZIÓŁKA, ZUCHNIARZ 2006). In agricultural practice, energetic plants can be grown on soils of different quality classes, successfully resisting the rising prices of food products (ROLA et al. 2007, GRZYBEK 2006). Willow plantations can be successfully set up on good soils – with quality classes I–III, but often soils of V and VI quality classes also meet the requirements of energetic plants. This is one of the ways of restoring and using lands, that are agriculturally degraded (KUŚ et al. 2008). From 1 ha of willow plantation one can get 20 tons of raw material with a calorific value equivalent to 10 tons of fine coal, thus producing wood in own farms saves the money spent on purchases of fuel, and in turn is equivalent to reducing the deficit in the household budget (GĄSIOREK 2005).

Cultivation of energetic plants is often placed on areas of great natural value, which will be lost as a result of the introduction of energetic crops. Growing energetic plants could take effect in a significant threat to biological diversity, because the area will be occupied in this way for 15 to 20 years (GRZYBEK 2006, NIEDZIÓŁKA, ZUCHNIARZ 2006). A characteristic feature of alternative crops is their pioneering and, therefore, not fully explored character. As indicated KORNIAK (2007), weed flora of willow is very similar in composition to the weed flora of northeastern part of Poland. It should be assumed though, that with the long cultivation of willow more specialized and burdensome weeds would develop. It should be remembered, however, that the combustion of willow (as an alternative to coal) helps reduce solid and gas wastes in energy production, so it has also a positive impact on the environment. In addition, willow, a plant with large and rapid increments, collects pollutants from the soil, enriching the environment with oxygen; it also regulates water relations (GĄSIOREK 2005, GRZYBEK 2006).

There remain many other issues to be resolved, such as yielding in climatic conditions of the individual regions, or the decision-making processes related to supply chain management (JANOWICZ, KUNIKOWSKI 2008). An important factor is also the seasoning of willow. Seasoning should be performed until the proper humidity is reached – about 25%, in the place of cultivation (FRĄCZEK, MUDRYK 2008, www.ieo.pl). It allows increasing the purchase price and the value of raw material.

Possibilities of *Salix viminalis* willow cultivation in eastern Poland voivodships

Eastern Poland regions are characterized by low population density in comparison with the rest of the country. The data presented in Table 1 indicate that the urban population equals to 46.59% of the total population, with a nationwide rate of 61.5%. At the same time, these voivodships account for nearly one third of the country area. These areas are then sparsely populated – the total number of population is 4.789 million, what is slightly less than one fifth of the country population (*Ludność według płci...* 2004). Different levels of socio-economic development in Poland occur due to geographical location, natural abundance of individual regions (*Efekty polityki spójności UE w Polsce*.

Table 1

Chosen indexes characterizing eastern Poland voivodships

Indexes	Voivodship			Poland
	Lubelskie	Podlaskie	Warmińsko-Mazurskie	
Population [tys.]	2,185.2	1,202.4	1,428.7	38,173.8
Rural population [tys.]	1,168.2	493.6	569.9	14,690
Total area [ha]	2,508.877	2,018.975	2,322.011	31,267.938
Area of fallow[tys. ha]	27.2	17.1	20.1	498.4
Percent of the fallow area in arable land	2.9	2.3	3.0	4.1
Consumption of electricity [GWh]	5,169	2,609	3261	134,473
Consumption of heat [TJ]	25,566	10,613	11,954	426,131

Source: own study based on GUS data from years 2004 and 2009.

2009) and the fact that the eastern voivodships of Poland are characterized by the lowest level of capital expenditures ”per capita” in the national economy in general (MIKOŁAJCZYK 2008).

The above considerations are inevitably reflected in the level of the eastern voivodships development when compared to the rest of the country. It should be noted though, that some of the features conditioning the slow development of the eastern Poland regions, also provide advantages in the context of the energetic crops introduction to this area. Among these features are:

- area of land not used for agricultural purposes,
- area of arable land with low soil quality class (i.e. IV, V and VI),
- sufficient human potential,
- appropriate science and research base.

The following subsections describe the specificities of the individual voivodships in the context of the above-mentioned conditions.

Warmińsko-mazurskie voivodship

The energy potential of energetic crops in the Warmia and Mazury voivodship is, according to PIECHOCKI and BIERANOWSKI (2007), equal to 47,250 TJ and is important in terms of energy balance of the area. Area of energetic crops currently cultivated in the region is about 300 acres and nearly the whole of this area is willow, while the forecast predicted that in Warmia and Mazury voivodship in 2009 will be grown 21 thousand ha of willow (*Program ekoenergetyczny województwa...* 2005).

Temporarily unused rural area in Warmia and Mazury voivodship was equal to 148.7 thousand ha in 1998 and 179.5 thousand ha in 2003. This amount is one of the largest, in comparison to other provinces (*Program ekoenergetyczny województwa...* 2005). If the human potential is added (such as high level of unemployment in rural areas), and appropriate science and research base (especially the University of Warmia and Mazury in Olsztyn), it appears that Warmia and Mazury voivodship has all the features predestening it for growing energetic plants.

Lubelskie voivodship

According to Biuro Planowania Przestrzennego in Lublin, willow cultivation is 2 times more profitable than growing wheat and six times more profitable than the cultivation of rye. This analysis, in contrast with the decline in profitability of agricultural production, results in a decrease in the income of farming families, and so it takes effect in noticable growth of establishments of new energetic plantations in Lubelskie voivodship. It should be noted that these crops are used primarily for farms own purposes.

Areas of fallow and idle lands in Lubelskie voivodship in 2004 was equal to 77.7 thousand ha, i.e. 6.5% of the total rural area of the region (*Wojewódzki program rozwoju alternatywnych...* 2006). There are large areas of arable land (about 152 thousand ha) and grassland (about 60 thousand ha) in the region, belonging to the IV, V and VI soil quality class (*Wojewódzki program rozwoju alternatywnych...* 2006). Therefore, this region has large reserves of land that could be used for energetic crops.

Podlaskie voivodship

Forecasts of energetic crops in Podlaskie voivodship developed by Polish Society for Biomass predict their steady growth. Podlaska Fundacja Rozwoju Regionalnego (*Praktyczne aspekty...* 2006) predicted that in Podlaskie could be grown 3 thousand ha of willow in 2006 and 5 thousand ha in 2009. Anticipated areas of crops in Podlaskie voivodship have not been met. As other voivodships, Podlaskie in the cultivation of energetic crops sees the possibility of reducing unemployment. Knowing its energy demand, the region is trying to introduce modern technologies involving refining biomass and creating biocarbon of high calorific value (about 28–32 MJ/kg) (*Praktyczne aspekty...* 2006). This action will allow the transfer of territory stigmatization and lengthening the distance from the farmer to the intended recipient.

Current consumption of energy obtained from energetic plantations in the Podlaskie voivodship is about 55 TJ, and estimates suggested further dynamic growth to 2010. It was anticipated that in 2010 the energetic plantations will result in 2,698 TJ of energy obtained annually. The energy potential, possible to be obtained from energetic plantations is, according to IBMER prognoses, not less than 17,940 TJ (*Praktyczne aspekty...* 2006).

Summary and conclusions

1. In Poland grows area of lands not used for agricultural purposes. Therefore, unquestionable alternative to the traditional agricultural production is the use of these terrains for energetic crops. The amount of such production should take into account environmental specifications and the expectations and limitations existing in specific region.

2. Coherence of legal provisions and actions relating to renewable energy, giving it priority status, will intensify the growth of this sector, eliminating the risks of not fulfilling obligations imposed on Poland by the European Union.

3. There is a possibility of energetic use of willow in both, micro (fuel), and macroscale (increasing share of RES in the country energy balance). The production of this type may increase the significance of eastern Poland regions and help in stopping the impoverishment of society in the region – this particularly applies to holders of soils of lesser quality.

Translated by SEWERYN LIPIŃSKI

Accepted for print 7.05.2012

References

BIENIEK J., ŻOŁNIERZ-RUSINEK A. 2008. *Wierzba salix viminalis jako źródło energii odnawialnej na przykładzie plantacji założonych na terenie kotliny kłodzkiej.* Inżynieria Rolnicza, 4(102): 111–118.

Charakterystyka obszarów wiejskich. 2008. Główny Urząd Statystyczny.

DAM J. VAN, FAAIJ A., LEWANDOWSKI I. 2005. *Biomass production potentials in central and Eastern Europe under different scenarios*. www.chem.uu.nl/nws/www/ publica/ Publicaties2005/E2005-87.pdf.

DENISIUK W. 2006. *Koszt likwidacji plantacji roślin energetycznych.* Inżynieria Rolnicza, 12(87): 99–107.

Dyrektywa 2001/77/EC z dnia 27 września 2001 o promocji energii elektrycznej wytwarzanej z odnawialnych źródeł energii na wewnętrznym rynku energii elektrycznej, s. 2–3.

Dyrektywa 2001/77/WE Parlamentu Europejskiego i Rady z dnia 27 września 2001 r. w sprawie wspierania produkcji na rynku wewnętrznym energii elektrycznej wytwarzanej ze źródeł odnawialnych.

Dyrektywa 2003/30/EC z dnia 8 maja 2003 r. w sprawie promocji wykorzystania biopaliw lub innych odnawialnych paliw w transporcie. Instytut Badań nad Gospodarką Rynkową, s. 145.

Dyrektywa 2003/87/EC z dnia 13 października 2003 r. regulująca sposób wypełniania zobowiązań Unii Europejskiej wobec Protokołu z Kioto, s. 2, pkt. 5.

Dyrektywa Parlamentu Europejskiego i Rady 2009/28/WE z dnia 23 kwietnia 2009 r. w sprawie promowania stosowania energii ze źródeł odnawialnych.

Efekty polityki spójności UE w Polsce. 2009. Dokument Problemowy Ministerstwo Rozwoju Regionalnego, Warszawa www.mrr.gov.pl/rozwoj_regionalny/poziom_miedzynarodowy/polityka_spojnosci_po_2013/Documents/Efekty_polityki_spojnosci_w_Polsce.pdf (acess: 3.05.2011).

Efekty polityki spójności UE w Polsce. 2009. Ministerstwo Rozwoju Regionalnego. http://www.mrr.gov.pl/rozwoj_regionalny/poziom_miedzynarodowy/polityka_spojnosci_po_2013/Documents/Efekty –polityki_spojnosci_w_Polsce.pdf (access: 3.05.2011).

Eurostat. 2008. *Environment and energy: Renewable energy statistic.*

FRĄCZEK J., MUDRYK K. 2008. *Zmiany wilgotności pędów wierzby salix viminalis w okresie sezonowania.* Inżynieria Rolnicza, 10(108): 55–61.

GĄSIOREK S. 2005. *Różnorodne aspekty upraw wierzby wiciowej w warunkach górskich.* Inżynieria Rolnicza, 6: 177–180.

Green Paper. 2000. Towards a European strategy for the security of energy supply COM/2000/0769 final.

GREWIŃSKI M. 2008. *Transformacja polityki społecznej-wyzwania na przyszłość w kontekście systemu pomocy społecznej.* http://www.mcps-efs.pl/do%20pobrania/inauguracja/7.pdf (access: 3.05.2011).

GRZYBEK A. 2006. *Wpływ wybranych roślin energetycznych na środowisko*. Problemy Inżynierii Rolniczej, 2: 81–88.

GRZYBEK A. 2007. *Analiza możliwości wykorzystania surowców rolnych na potrzeby produkcji biopaliwa.* Problemy Inżynierii Rolniczej, 1: 163–169.

JANOWICZ L., KUNIKOWSKI G. 2008. *Ocena zasobów odnawialnych w oparciu o system informacji geograficznej (GIS).* Inżynieria Rolnicza, 4(102): 329–335.

KORNIAK T. 2007. *Zachwaszczenie upraw wierzby w północno-wschodniej części Polski*. Pamiętnik Puławski, 145: 141–149.

Kukuła S., Krasowicz S. 2006. *Regionalne zróżnicowanie polskiego rolnictwa w świetle badań.* IUNG-PIB Puławy.

Kuś J., Faber A., Stasiak M., Kawalec A. 2008. *Plonowanie wybranych gatunków roślin uprawianych na cele energetyczne na różnych glebach*. Problemy Inżynierii Rolniczej, 1: 79–86.

Kwaśniewski D. 2008. *Produkcja biomasy a koszty surowcowo-materiałowe na jednorocznych plantacjach wierzby energetycznej*. Inżynieria Rolnicza, 10(108): 153–158.

Lesz M. 2006. *Odnawialne Źródła Energii w Polsce – dokumenty programowe i wsparcie finansowe.* Materiały na konferencję „Realizacja inwestycji biomasowych – aspekty praktyczne".

*Ludność stan i struktura w przekroju terytorialnym.*2009. Główny Urząd Statystyczny.

Ludność według płci, wieku, województw, podregionów powiatów, miast i gmin. 2004. Główny Urząd Statystyczny.

Mikołajczyk J. 2008. *Regionalne zróżnicowanie wydatków inwestycyjnych w rolnictwie polskim w latach 2000–2005.* Stowarzyszenie ekonomistów rolnictwa i agrobiznesu. Roczniki Naukowe, X(2).

Myszczyszyn J. 2001. *Problemy bezrobocia ukrytego w rolnictwie chłopskim. Przemiany i perspektywy polityki gospodarczej.* SGH, Warszawa.

Niedziółka I., Zuchniarz A. 2006. *Analiza energetyczna wybranych rodzajów biomasy pochodzenia roślinnego* MOTROL 8A: 232–237.

Ocena stanu i perspektyw produkcji krajowej urządzeń dla energetyki odnawialnej. http://www.ieo.pl/pl/ekspertyzy.html (access: 3.05.2011).

Pawlak J. 2007. *Udział energii z zasobów odnawialnych w gospodarce narodowej i w rolnictwie.* Problemy Inżynierii Rolniczej, 1: 87–93.

Pawlak J. 2009. *Przewidywane zmiany w mechanizacji produkcji roślinnej w Polsce do roku 2020.* Studia i raporty IUNG-PIB, 14: 328–340.

Piechocki J. Bieranowski J. 2007. *Metodyka programu ekoenergetycznego dla wybranego obszaru administracyjnego z uwzględnieniem odnawialnych zasobów energii pochodzenia rolniczego.* Inżynieria Rolnicza, 8(96): 21–27.

Praktyczne aspekty wykorzystania odnawialnych źródeł energii plan energetyczny województwa podlaskiego. 2006. Podlaska Fundacja Rozwoju Regionalnego. http://paze.pfrr.pl/pliki/Praktyczne_aspekty_wykorzystania_odnawialnych_zrodel_energii.pdf (access: 3.05.2011).

Prognoza rozwoju rynku odnawialnej energetyki elektrycznej do roku 2020, z uwzględnieniem perspektywy roku 2030. 2009. Fundacja na rzecz energetyki zrównoważonej. http://www.fnez.org/upload/file/24.pdf (access 3.05.2011).

Program ekoenergetyczny województwa warmińsko-mazurskiego na lata 2005–2010. 2005. Zarząd Województwa Warmińsko-Mazurskiego. http://www.wmae.pl/userfiles/file/Do%20pobrania/program%20ekoenergetyczny%20woj%20warm_maz%202005_2010.pdf (access: 3.05.2011).

Protokół z Kioto zobowiązujący kraje rozwinięte do redukcji emisji gazów cieplarnianych do atmosfery w latach 2008–2012, s. 2, 3, art. 2, 3. Dz.U. z dnia 17 października 2005 r.

Rola J., Sekułowski T, Rola H., Badowski M. 2007. *Bioróżnorodność zbiorowisk chwastów na plantacjach wierzby krzewiastej (salix viminalis) na terenie województwa dolnośląskiego i opolskiego.* Pamiętnik Puławski, 145: 1–11.

Roszkowski A. 2008. *Biomasa kontra rolnictwo.* Inżynieria Rolnicza, 10(108): 201–208.

Siejka K., Tańczuk M., Trinczek K. 2008. *Koncepcja szacowania potencjału energetycznego biomasy na przykładzie wybranej gminy województwa opolskiego*. Inżynieria Rolnicza, 6(104): 167–174.

Sołowiej P. 2006. *Możliwości zastosowania odnawialnych źródeł energii w wybranym indywidualnym gospodarstwie rolnym.* Inżynieria Rolnicza, 11(86): 447–454.

Szwedziak K. 2010. *Energetyczne wykorzystanie słomy na terenie województwa opolskiego.* Inżynieria Rolnicza, 5(123): 275–281.

Ścibisz M. 2006. *Możliwości wykorzystania energii słonecznej w gospodarstwach województwa lubelskiego.* Inżynieria Rolnicza, 13(88): 437–442.

Walski J. 2007 *Raport o stanie energetycznym gmin w Polsce.* http://www.preda.pl/pliki/Dokumenty/Komentarze/Raport_o_energetyce_sieciowej.pdf (access: 3.05.2011).

Wojewódzki program rozwoju alternatywnych źródeł energii dla województwa lubelskiego. 2006. Biuro Planowania Przestrzennego w Lublinie. http://www.oze.bpp.lublin.pl/dokumenty/program/prognoza.pdf (acess: 3.05.2011).

Województwa w latach 1995–2005. 2005. Główny Urząd Statystyczny.

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

ENERGY AND EXERGY FLOW BALANCES FOR TRADITIONAL AND PASSIVE DETACHED HOUSES

Zygmunt Wierciński, Aldona Skotnicka-Siepsiak

Chair of Environmental Engineering
University of Warmia and Mazury in Olsztyn

Key words: energy conversion in building, exergy balance, passive house.

Abstract

The aim of this paper is to give the insight into the energy and exergy analysis and the usefulness of the exergy balance next to the energy balance for the evaluation of the different kind of buildings. In our case we applied this method to the traditional and low-energy (passive) showing the differences in heating systems used for these buildings. In traditional house the condensing boiler and water heating radiator were used while in the passive the heat pump and air heating were used. The exergy analysis showed that the exergy destruction for the low-energy house is much lower than for the conventional one.

BILANS ENERGII I EGZERGII DLA DOMU JEDNORODZINNEGO TRADYCYJNEGO I PASYWNEGO

Zygmunt Wierciński, Aldona Skotnicka-Siepsiak

Katedra Inżynierii Środowiska
Uniwersytet Warmińsko-Mazurski w Olsztynie

Słowa kluczowe: przemiana energii w budynku, bilans egzergii, dom pasywny.

Abstrakt

Celem artykułu oprócz bilansu energii, jest bilans egzergii oraz wykazanie jej przydatności do oceny różnych rodzajów budynków. Zastosowano tę metodę do budynku tradycyjnego i niskoenergetycznego (pasywnego) i wskazano na różnice w zastosowaniu różnych systemów ogrzewania w budynkach. W budynku tradycyjnym zastosowano kocioł kondensacyjny i grzejniki radiacyjne, w budynku pasywnym pompę ciepła i ogrzewanie powietrzne. Po analizie egzergii wykazano, że w budynku niskoenergetycznym destrukcja egzergii jest znacznie mniejsza niż w budynku konwencjonalnym.

Introduction

Drastically decreasing resources of fossil fuels cause the need to search for the energy efficient technological solutions. In the field of housing technology and construction those trends have resulted in the designs of the so called low-energy and passive houses. Those, however, are not ultimate and fully satisfying solutions (exhausting the potential of scientific and technological development). Further reducing of heat demand for buildings heating will probably require application of new notions and analytical tools allowing the energy efficiency assessment of various heating circuits. This assessment will be done to answer the question where in the entire process of energy flow and conversion the exergy losses are the greatest and occur the most frequently. Energy and additionally exergy analysis with respect to the building could give us such a tool for analysis the heat demand and the choice of proper solution. Actually, this analysis is only partially, because it does not take into account the analysis of exergy in the different processes of production of building materials and components.

The goal of the paper is to show the suitability of exergy analysis – made beside and not instead of the energy analysis – to present the actual processes of energy conversion during its flow (use) in residential buildings in view of computations and to show it on an example of the comparison of two detached houses: the traditional with the passive.

Idea of passive house versus traditional

Exergy loss during the energy conversion (the notion of energy consumption or even worse energy loss is very often used – not only in common language – to describe this process, but it is rather not appropriate designation) for heating the building covers three types of needs: (1) exergy lost through external partitions of the building as the result of the processes of heat conductivity, conversion in wall and radiation, (2) exergy lost as a result of exchange of air in the building for the purpose of ventilation and (3) exergy lost in the energy flow needed to heat the domestic hot water for general use. Continuous progress in technology of building materials, in manufacturing of windows and doors allow decreasing utilization of energy to satisfy the needs of the first type. Application of ground heat exchangers and heat recovers in blow in-blow out ventilation systems decreases the needs of the second type. Solar energy is used to heat water increasingly frequently.

Thanks to the technological development within the last 40 years the demand for energy to heat the buildings decreased by almost a half (from 80%

to 45%) while energy utilization related to ventilation and water heating remained practically unchanged because of the need to secure or enhance the appropriate hygienic conditions. As a consequence, in the energy balance of housing of 1990s the percentage share of energy utilization for ventilation and domestic water heating is relatively high at 30% and 25% respectively.

Shrinking resources of fossil fuels motivate the search for solutions that would allow further limitation of energy utilization during use of residential buildings. Implementation of the notion of low energy house and attempts at practical implementation of that idea are manifestations of those efforts. Low-energy house should consume 30% less energy as compared to a traditional house. Energy consumption per one square meter per year in the low-energy house with the usable area of 150 square meters should not exceed 35 kWh for heating purposes, 35 kWh for ventilation of the space and 15 kWh for water heating. The additional decrease in fossil fuels consumption for satisfying those needs is possible thanks to, among others, application of condensation technology or heat pumps in the building heating installations.

Further attempts at decreasing energy consumption resulted in the idea of the passive house. The notion of the passive house designates an object – roughly speaking – that does not require the active heating system during the house operation. In such a house that significant reduction in heat losses is possible through application of maximally insulated external partitions and tight windows with triple glazing, passive use of solar energy and controlled ventilation with recovery of heat from the air blown out. The so called passive sources in the form of solar energy passing through the windows, heat generated by the residents and heat generated from the use of household equipment especially electric devices is sufficient to cover the exergy losses during the heating. Only during the period of decreased temperatures supplementary (usually ecologically friendly) heating is used, usually based on the heat supplied by the ventilation installation. In the area of Central Europe the passive building have to be equipped with mechanical blow in-blow out ventilation with heat recovery with maximum insulation of the partitions. Appropriate performance of all the above improvements results in obtaining the building offering high heat comfort and very low demand for heat energy not exceeding 15 kWh/ $(m^2 \cdot y)$.

Despite the obtaining of such low heat-demands of the passive house for energy as compared to traditional construction further attempts at minimizing energy outlays are undertaken. It seems that energy balance based on the first principle of thermodynamics is not the sufficient analytical tool in that field. That balance treats in the same way the different forms of energy without considering their different quality and practical value. On those bases it is difficult to asses the energy efficiency of heat circuits and answer the question where in the entire process of the flow the largest losses take place and, as a consequence, in which link further improvements should be searched for. The need appeared to introduce a thermodynamic notion of exergy necessary to characterize the problems considered. The notion of exergy is actually not a new one, but its application for the building and its service systems is rather new.

A few words on exergy

The notion of exergy means the maximum ability to perform work by a given matter determined by considering the participation of environment. Exergy can serve determining the practical use of energy contained in matter. The assessment of that use requires knowledge of total quantity of exergy of the considered thermodynamic medium and the relation of exergy to the volume of the medium. That parameter is called the exergy density.

One of the most important differences between the notion of energy and exergy is that energy is subject to the conservation law while the exergy conservation law does not exist. Additionally, energy has conventional reference levels while in case of exergy the reference levels are imposed by the nature ie. environment. Losses of exergy are inevitable and at the same time unwanted as every loss of exergy cause a decrease of the use effect of a given process (SZARGUT, PETELA 1965, SZARGUT 2005, WALL 1999). The investigations by Gouy – Stodola indicate that loss of exergy is irreversible and cannot be recovered even partly. That characteristic of irreversible loss of exergy differentiates it from the losses of work determined by other methods. In case of investigation on exergy the mutual relation of phenomena occurring in individual links of the compound process should be considered. If, in the process of investigation it is found out that the loss of exergy is particularly substantial in a given link of the compound process than possibilities of decreasing the level of irreversibility of transformations within that link should be investigated and possibilities of changes in the progress of processes at earlier stages should be analyzed.

Losses of exergy can be divided into internal and external. Loss caused by irreversible transformations occurring within a considered device are called the internal loss. The causes of internal losses of exergy can include, among others, heat flow at finite temperature difference, transfer of work coupled with lack of mechanical balance, mixing of substances possessing different chemical compositions or loss of liquid pressure during flow caused by viscosity.

On the other hand, destruction of exergy caused by irreversible equalization of the parameters of waste media with the parameters of environment media is defined as an external loss. In the majority of energetic thermodynamic processes discharge of waste thermodynamic medium into the environment (e.g. discharge of flue gas from the boiler) takes place. Parameters of the components of the discharged medium differ generally from the parameters of the environment. The waste medium still possesses a certain unused exergy. In the environment the destruction of this exergy takes place.

As a consequence of the mentioned exergy losses the exergy does not satisfied a low of conservation. The difference between input and output exergy is equal to the internal exergy loss. As a consequence the exergy balance can be represented by the following formula:

$$B_{\mathrm{d}} = \Delta B_{\mathrm{u}} + B_{\mathrm{wus}} + L + \Sigma\Delta B_{\mathrm{sr}} + \delta B_{\mathrm{w}} + \delta B_{\mathrm{z}} \tag{1}$$

where:

B_{d} – exergy of system input substances,
ΔB_{u} – increase of system exergy,
B_{wus} – useful exergy of useful input products,
L – mechanical or electrical work done by the system,
$\Sigma\Delta B_{\mathrm{sr}}$ – increase of exergy of the external heat source operating in the control casing of the system,
δB_{w}, δB_{z}– internal and external exergy loss respectively.

According to the simplest representation the exergy balance equation can take a following form:

$$B_{\mathrm{i}} - B_{\mathrm{cons}} = B_{\mathrm{o}} \tag{2}$$

$$B_{\mathrm{cons}} = T_{\mathrm{ref}} \cdot S_{\mathrm{gen}} \tag{3}$$

Where B_{i} and B_{o} are exergies at the input and output of the system and B_{cons} is the exergy destroyed (consumed) in the system, which in turn is equal to the product of reference temperature T_{ref} and the entropy S_{gen} produced in the system. Usually the environmental temperature is acknowledged as a reference temperature in the building analysis of exergy.

Reference to the notion of entropy in that balance allows explaining what can be used and what can be generated. The notion of exergy is used to express what is used, consumed, while entropy describes what is removed. Exergy means the ability of energy to dissipate during its flow through the system while entropy expresses the state of dissipation. In the context of the purpose of this paper it is important that the theory of exergy can be applicable in assessment of reversibility and efficiency of all energetic processes, in which heat radiation participates that is in case of devices using energy of solar radiation and heating system operating through radiation.

The most important goal of exergy analysis is detecting and quantitative assessment of causes of the thermodynamic imperfection of thermal processes and it can contribute to improvement of those processes. That analysis can be highly useful as a tool in research works aiming at increasing energy efficiency of housing construction as concerns maintaining the conditions of residents thermal comfort with possibly low energy utilization. The aim of such works is to design objects and heating systems with possibly low exergy (i.e. decreasing the input of exergy during heat generation and management). Low exergy of the heating system encompasses: the increased level of use of the chemical energy of fuels (or substitution of fossil fuels with devices using renewable energy sources), low consumption of heat in the building and use of low temperature heat for heating purposes. All that should result in decreasing the consumption of fossil fuels and decrease of the emissions of pollutions, production of waste and environmental losses.

Energy and exergy balance spreadsheet of Annex 49

Development within Annex 37 software of the spreadsheet for exergy balance analysis was one of the effects of works on decreasing the residential buildings heating exergy (Collective work, 2003). It is further complimented and improved in Annex49 (Torio, Schmidt 2011) to the version 7.7.

The spreadsheet is divided into seven modules. In the first of them the basic data concerning the building such as its volume, net area, external and internal temperature is input. The second module covers the heat losses involved in heat transmission. It covers heat losses through partitions (on the basis of the area of partitions and their heat transmission coefficients as well as the difference between the external and internal temperatures) as well as loses for ventilation (computer, among others on the basis of air infiltration, exchanger efficiency and temperature difference). Parts three and four determine the possible heat gains. They include gains from solar radiation (taking into account the surface of windows and their orientation in relation to the sides of the world) and internal gains (in the form of residents; body heat and heat generated by household appliances). In module five the losses and gains are balanced. In part six the types of installations and energy source should be chosen. On the basis of the choice of the heat generating system the spreadsheet describes the efficiency, maximum supply temperature and the degree to which the system uses the renewable energy sources using “macros”. Similarly, in selecting the heat generating system the spreadsheet selects characteristic values for a given type of installation such as the supply and return temperature, output and additional energy. Exergy analysis is done in the last module of the code. It progresses describing one by one the heat flow stages: building envelope (walls, roofing), air in premises, heat discharged, heat distribution, container, generation, basic transformations (Fig. 1). In Figure 1, apart of the mean stream of energy the additional stream of auxiliary energy in the form of electricity is shown.

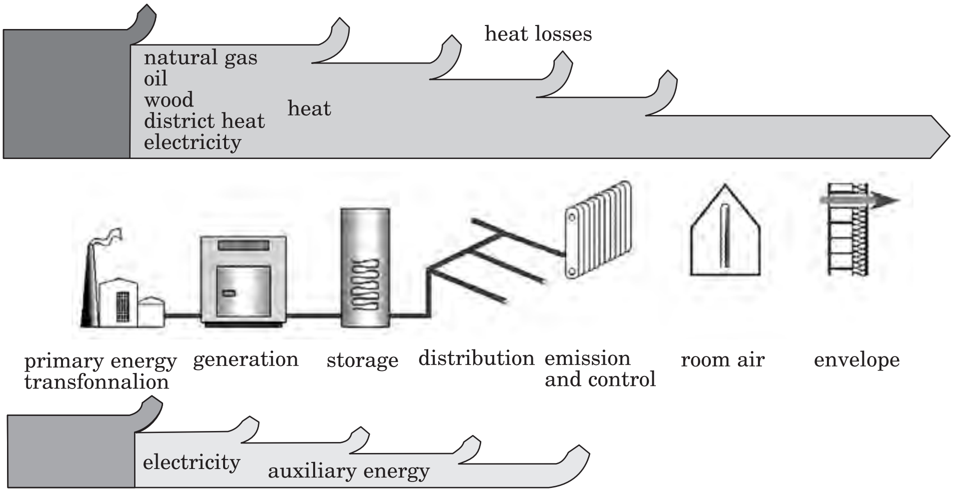


Fig. 1. Energy stream in building services

Source: Schmidt (2004).

In the exergy calculation of the room air only the thermal part of the heating exergy is taken into account and the changes of the exergy connected with the change the humidity and pressure of the air both inside and outside of the building are not considered. Sakulpipatsin, (2008), has calculated the exergy changes for three different climates (hot and wet as in Bangkok, rather cold in Holland and finally sea climate in Lissabon) and he came to the conclusion that the error caused by the neglecting of the wet part of exergy for the cold climate can caused not more than 3% error in exergy calculations.

Heat transmission through partitions

We start the computation with determining the thermal load of the building considering transmission coefficients of the window, walls, doors and others according to the formula beneath:

$$\Phi_{tr} = \Sigma(U_i \cdot A_i \cdot F_{xi}) \cdot (\Theta_i - \Theta_e) \qquad (4)$$

where:

A_i – area of i-th partition in [m^2] (walls, windows, doors and others),

U_i – heat transmission coefficient for the *i*-th partition in [W/m^2 K],

Θ_i, Θ_e – indoor and exterior air temperatures under the design conditions in [K],

$F_{x,I}$ – temperature correction coefficient.

Heat loss for ventilation

Next the heat flow connected with the ventilation air is considered along the following formula:

$$\Phi_{V,is} = (c_p \cdot \rho \cdot V \cdot n_d \cdot (1 - \eta_V)) \cdot (\Theta_i - \Theta_e) \qquad (5)$$

c_p – specific heat of air in [kJ/kg deg],

ρ_a – density of air in [kg/m^3],

V – building volume (capacity) in [m^3],

n_d – air exchanged rate (ach) in [1/h],

η_V – efficiency coefficient of heat exchanger (recuperator) if used.

Solar heat gains Φ_S

The sun radiation gains in the building can be calculated with the formula below:

$$\Phi_{sol,gn} = \Sigma(I_{sol,j} \cdot (1 - F_{fr}) \cdot A_{Wj} \cdot g_j) \qquad (6)$$

where:

$I_{sol,j}$ – intensity of solar radiation (dependent on window orientation to world directions and slope angle to direction of radiation) [W/m^2],

F_{fr} – window frame fraction, constant for all windows used in building [–],

A_{wj} – area of *j*-th window [m^2],

g_j – total transmittance of *j*-th window [–].

Internal heat gains Φ_i [W]

These heat gains can be divided into two different parts:
– heat gains from occupants:

$$\Phi_{in,gn,o} = no_o \cdot \Phi''_{in,gn,o} \tag{7}$$

where:
no_o – number of occupants,
$\Phi''_{in,gn,o}$ – heat gains from one person [W]

– Heat gains from internal devices

$$\Phi_{in,gn,equ} = \Phi''_{in,gn,equ} \cdot A_{ni} \tag{8}$$

where:
$\Phi''_{in,gn,equ}$ – specific heat gains from devices [W/m^2],
A_{nt} – netto floor area [m^2].

Another heat gains

Finally two different heat gains are considered especially from lighting and electromotors especially driving the fan:
– lighting [W]

$$P_L = p_L \cdot A_{nt} = \Phi_{in.L} \tag{9}$$

where:
p_L – specific power of lighting [W/m^2],
A_{nt}– netto floor area [m^2].

– gains from the fan motor [W]

$$P_V = p_V \cdot V \cdot n_d \tag{10}$$

where:
p_V – specific fan motor power in [Wh/m^3],
V – volume of building [m^3],
n_d – air infiltration, [ach/h].

Heat demand

At the end of calculation the heat demand and its specific value for the building can be given by following formulae:

$$\Phi_H = (\Phi_{tr} + \Phi_{V,is}) - (\Phi_{sol,gn} + \Phi_{in,gn,o} + \Phi_{in,gn,equ} + \Phi_{in,L}) \tag{11}$$

And also the specific heat demand:

$$\Phi''_H = \frac{\Phi_H}{A_{nt}} \tag{12}$$

Exergy load calculations

In this part of our paper we will show some formulae relating to the exergy calculation used in the code Exergy 7.7.

At the beginning it is necessary to consider the exergy loss understood as the exergy flow through the envelope of the building. So it the exergy needed to sustain the temperature difference between the interior and exterior condition of the building. Thus, the value of exergy for heating the premises results from the formula:

$$B_{heating} = F_{q,room}\Phi_H \tag{13}$$

where

$$F_{q,room} = (1 - \frac{T_e}{T_{in}}) \tag{14}$$

is called a quality factor of room air, and it is in the form of the so called Carnot coefficient, and where T_e and T_{in} are the external environmental and internal temperature of the air. In the case of T_e = -22°C and T_{in} = 20° the coefficient $F_{q,room}$ = 0.14.

When the air in the building is heated by the heater (radiator or floor, or any other) then next the exergy load for the heater should be calculated. So next the different temperatures of the heater should be defined as follows: T_{ing} and T_{ret} as the temperature of the inlet and outlet of the heater, and T_{in} is the temperature in the heated room or in whole building.

First, the mean temperature of the radiator T_{heater} can be calculated according to two following formulae:

$$T_{heater} = 0.5 \cdot T_{Ln} + T_{in} \tag{15}$$

Whereas T_{Ln} is the averaged logarithmic temperature of radiator:

$$T_{Ln} = \frac{(T_{ing} - T_{in}) - (T_{ret} - T_{in})}{\ln(T_{ing} - T_{in}) - \ln(T_{ret} - T_{in})} \tag{16}$$

Then the exergy stream from the radiator (heater) will be given by:

$$B_{heater} = \Phi_{heater} \cdot F_{q,heater} \tag{17}$$

Where $F_{q,heater}$ is the quality factor of air at heater (Carnot coefficient) given by below formula:

$$F_{q,heater} = 1 - \frac{T_e}{T_{heater}} \tag{18}$$

Of course, the heater belongs to the so called emission part of the system, Fig. 1. It is necessary to consider the loss in it and to calculate the necessary demand of energy and exergy caused by the loss in this part of the system.

The additional demand for the heating energy should be calculated according to following formula:

$$\Phi_{ls,em} = \Phi_H \cdot \left(\frac{1}{\eta_{em}} - 1\right) \tag{19}$$

Where η_{em} is the efficiency of the emission system.
The exergy demand for the emission system is to be calculated along formula given below:

$$B_{em} = \Delta B_{em} + B_{heat} = \left\{\frac{(\Phi_H + \Phi_{ls,em})}{(T_{in} - T_{ret})}\right\} \cdot \left\{(T_{in} - T_{ret}) - T_{ref} \cdot \ln\left(\frac{T_{in}}{T_{ret}}\right)\right\} \tag{20}$$

Where T_{ref} is usually the environmental temperature of air, as mentioned above.

In very similar way the exergy and energy load in the distribution system and storage can be calculated only using another efficiency factors and temperatures at the inlet and return. The reference temperature remains the same.

It is also possible by means of the Exergy code version 7.7 to calculate the energy and exergy demand according to the number of occupants and amount of DHW for one person per day according to the appropriate standard. And this possibility was used in our calculations of energy and exergy use in the detached family houses.

In each part of the system some additional energy is needed to drive the system usually in the form of electrical power. This auxiliary energy and its exergy should be taken into account.

And finally at the begin of the chain, Figure 1, stands the generation part of the system energy and primary energy transformation usually in the form of chemical energy and exergy of the fuel. To take into account this part of the system the efficiency of the generation (e.g the efficiency of boiler or heat pump) of energy and exergy and its auxiliary energy load is to be included into the balance. If the primary energy transformation is considered and the renewable energy, as the solar gains, Eq. 6, is taken into account the total energy and exergy amount are the results. Actually, all calculation of the energy demand for heating of the building are made according to the European standard EN-13790 Energy performance of buildings. Calculation of energy for space heating and cooling.

Description of houses under consideration

The computations were made for a passive house and traditional technology house according to the design by the Architecture Office Lipinscy Houses. In can be supposed that the first detached passive house designed by the Lipinscy Domy was built in Smolec near Wrocław, see also (Lipińscy Domy).

For the purpose of our calculations the both houses are placed near Olsztyn so in the fourth climatic zone with the standard winter temperature (average) in January -22°C. Both houses have the same heated volume of 415.9 m^3 and heated area of 142.3 m^2.

Passive house

The design was developed in collaboration with the specialists from the Institute of Passive Buildings at the Polish National Energy Conservation Agency in Warsaw. It received the positive opinion of the Passive House-Institute from Darmstadt and at the further stage certification by that

Institute. According to the energetic certificate issued by the above institutions its demand for thermal power is to be 1.9 kW (13.5 W/m^2), year final energy (heat) consumption for heating the building 13.5 kWh/m^2/year and 4.7 kWh/m^3/year. The seasonal demand for heat to heat that building will be 1944 kWh/year. This data are for the passive house located near Wrocław.

As mentioned above, the construction of the passive house is made according to the passive house regulations, and especially the thermal transmittance U of the building elements are appropriately chosen: exterior wall thermal transmittance is equal to 0.1 W/(m^{2}K), and appropriately the thermal transmittance of window 0. 60 W/(m^{2}K), door 0.80 W/(m^{2}K), roof 0.15 W/(m^{2}K), and finally floors to ground 0.11 W/(m^{2}K). So the superinsulation was employed to lower significantly the heat conductivity through the house envelope.

The insulation layers are 30–44 cm thick. The house is made of gravelite-concrete prefabricated walls and the thermal bridges are actually minimized as often as possible.

This house is designed fulfilling the passive solar design technique: it is possibly compact in shape to reduce its surface area with the shape coefficient equal to 0.75, the windows are oriented south to maximize the solar gains, but the solar gains are of secondary importance in minimizing the total energy demand requirement. The windows of advanced technology was applied with possible small U = 0.8 W/(m^{2}K), for the entire window including also the frame. So in the house the door of Clima Design and window woodwork of REHAU Company were used. Of course the airtightness of walls are very important in the passive house, and the special test for tigthness against the air penetration was necessary giving the result 0.3 air change per hour.

The Lipinscy passive house is equipped with the mechanical ventilation including the heat recovery system. According to the passive house standards the ground heat exchanger was provided for and it is Awaduct Thermo of the REHAU Company. The ground heat exchanger heating initially the air supplying the heat pump (air-water) and the building was also installed.

In case of the passive house the temperature of air blown in past the ground heat exchanger was assumed to be 5°C. In this case electric heating installation and hot water preparation was applied using the integrated compact Vitotres 343 device by Viessmann (air-air heat pump, heat recovery device and electrically heated container of water for general use). The heat pump in the heat exchanger heats the air blown into the room after leaving the heat recovery device. To prepare the DHW the vacuum solar heat collectors Vitosol of Viessman were additionally used.

In Figure 2 the scheme of energy streams (with appropriate temperatures) are shown for the passive house heating. The solar collector used for DHW preparation is not shown in this figure.

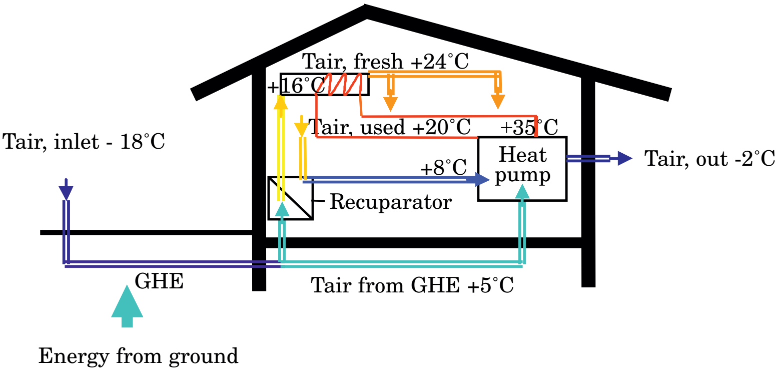


Fig. 2. Scheme of the heating system for the passive house

Traditional house

The traditional house was supposed to be made on the basis of a similar design as the passive house and had the same heated space of volume of 456.1 m^3 and heated area of 153.7 m^2. The difference in construction lies in the thermal transmittance of external partitions made of different materials than the passive house. Thus, especially, the thermal transmittance U of the building elements are appropriately changed: exterior wall thermal transmittance is equal to 0.37 W/(m^{2}K), and appropriately the thermal transmittance of window 1.0 W/(m^{2}K), door 1.4 W/(m^{2}K), roof 0.27 W/(m^{2}K), and finally floors to ground 0.22 W/(m^{2}K).

The heating installation in the conventional house consisted of gas condensation boiler and central water heating installation equipped with plate conventional heaters. The boiler also served heating the water. The natural ventilation is only supposed.

Results of calculations

In what follows, the results of computations made using the spreadsheet designed within the frameworks of IEA ECBCS Annex 37 and improved and complemented in the Annex49 are presented. The results are shown in Table 1 and 2. It is easy to see that the heat losses through partition are about 80 percent higher for the traditional house than for the passive one. Furthermore the loss for ventilation is five times higher for the traditional house than passive, what is quite clearly because there is natural ventilation in traditional house versus mechanical in passive house. The mechanical ventilation is also seen in the line for the electrical consumption of electricity for ventilator, where is nil for the traditional house versus 112 W in the passive house. Again the gains from the solar radiation in the case of passive house is 2.5 times higher than in traditional house. Three remaining gains: internal from resident and from electric devices and from lighting solar are the same because of the same number of inhabitants (four persons) and the number of electric devices and the power of lighting.

Table 1

Summary of the demand for thermal power of the traditional and passive house

Components in [W]	Traditional house	Passive house
Heat losses through partitions	5,531	3,045
Heat losses for ventilation	3,511	702
Gains from solar radiation	108	262
Internal heat gains (from residents)	320	320
Internal heat gains (from electric devices)	270	270
Electricity consumption for lighting	284	284
Electricity consumption for ventilators	0	112
Demand for thermal power	8,059	2,610
Demand for thermal power per 1 m^2 of usable area	56.6	18.3

Table 2

Results of exergy calculation

Components in [W]	Traditional house	Passive house
Exergy load room (envelope)	1,154	373
Exergy load at heater (room air)	1,616	408
Exergy load emission	2,134	471
Exergy load distribution	2,371	547
Exergy load storage	2,472	580
Exergy load generation	10,874	752
Exergy load transformation	11,364	2,258
Lighting exergy demand	284	284
Ventilation exergy demand	0	112
Exergy load plant	433	461
DHW prim exergy demand	471	153
Energy/Exergy prim load plant	1,067	1,134
Primary energy renewable	13	629

Finally, the demand for thermal power in traditional house is equal to circa 8.1 kW, when in the passive house is only 2.6 kW. And this huge difference is actually caused by two things: much better isolation and mechanical ventilation in the case of the passive house. Actually, the solar gains in the case of passive house could be a bit higher because it was impossible to calculate it in the spreadsheet Exergy ver. 7.7 according to the rules given for the passive house. The solar gains estimation along the rules for the passive house looks also to some extend optimistic. The specific power for the traditional house is equal to 56.6 W/m^2 and is more than three times higher than for the passive house, which is equal to 18.3 W/m^2.

In Table 2 the results of exergy flow calculations are quoted. These results are more instructive, because the exergy flow points out how much exergy finally disappears.

In Figure 3 and 4 the comparison of energy calculation for the conventional and low-energy house are illustrated. In Figure 3 the energy gains and losses are shown in bar diagram for the conventional, while in Figure 4 the same calculation for the low-energy house (passive) are given. The differences are very striking: the energy losses in passive house are 50% lower than in traditional house. In passive house the renewable energy gains (green) are rather huge, while in traditional house are equal to zero. The total input for the traditional house is equal to 13.525 kW, while for the low energy house is only 5.5 kW. The solar gains in traditional house are small in comparison to passive house, the internal gains are the same the same, because of the same number

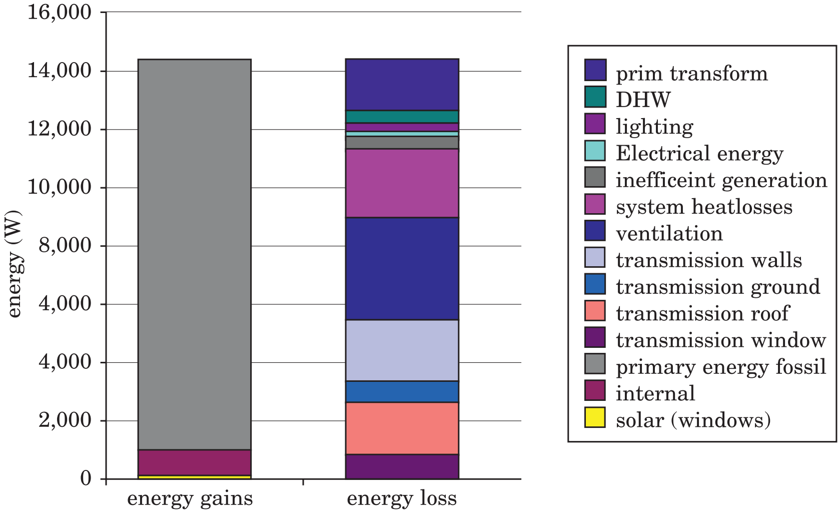


Fig. 3. Energy gains and losses in the traditional house

of occupants and lighting gains. Very instructive is also the comparison of exergy supply and demand for the traditional and low-energy house shown in Figures 5 and 6 respectively. The exergy supply and demand in the case of low-energy house is about two and half times lower than in the traditional house. It is the proper measure what exergy (energy) savings are achieved in the case of the low energy (passive) house.

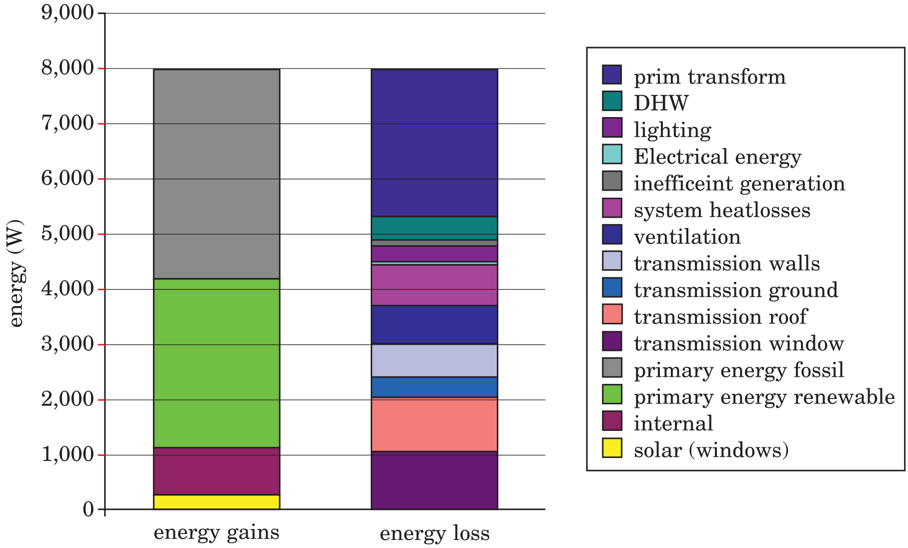


Fig. 4. Energy gains and losses in the low-energy (passive) house

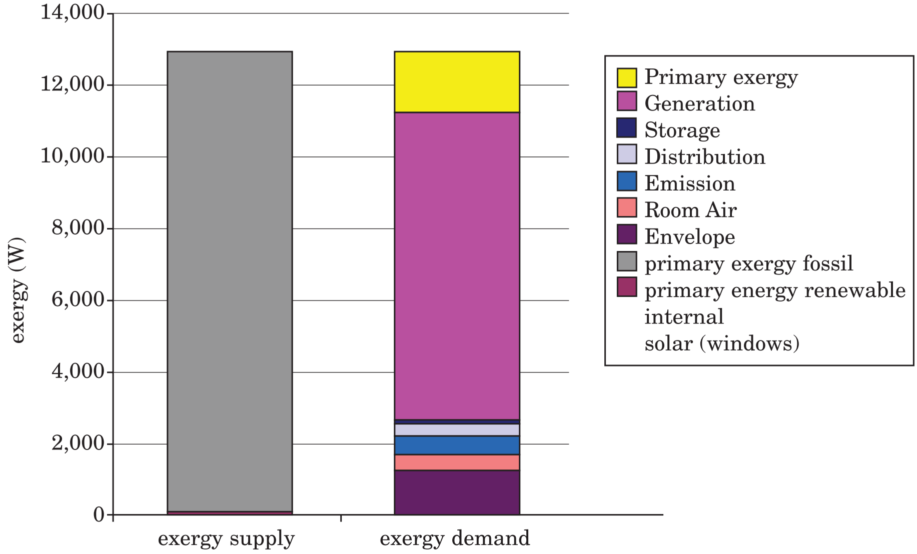


Fig. 5. Exergy supply and demand in traditional house

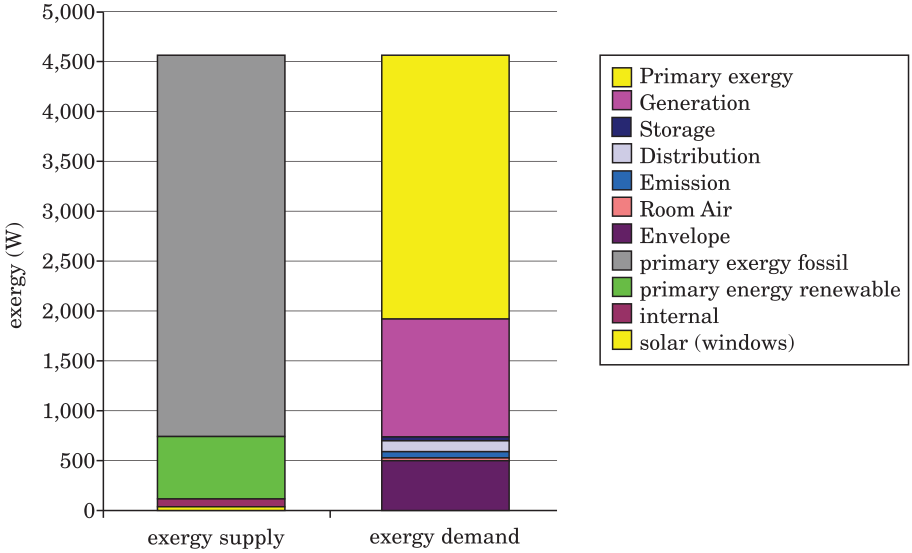


Fig. 6. Exergy supply and demand in low-energy (passive) house

Conclusions

The energy and exergy analysis of buildings are introduced according to the Energy Conservation for Building and Community Systems (ECBCS) Annex 49 and 37 of International Energy Agency. The code Exergy 7.7 was used to accomplish the energy and exergy flow analysis for the traditional and low-energy houses. It was shown that total demand for thermal power (converted to primary energy) of the traditional house was almost two times higher than that of the passive house. On the other hand losses of exergy at heat generation for central heating stage in the first stage were almost eight times higher than the losses of exergy in the passive house. The highest exergy losses occurred in the heat source (particularly the water boiler).

Large exergy losses also occur in heat transmission (distribution) to heaters (or in air channels to blowers in the passive house). In case of the traditional house losses of exergy were five times higher as a consequence of the difference in supply temperatures of water 70°C and air 35°C systems. As a consequence there were also around 3.5 times higher losses of exergy in transmission of heat from heating elements to the air in the premises.

Poorer insulation properties of the traditional building caused only two times increase of losses of exergy through the roofing (as compared to seven times higher losses in the boiler).

The exergy analysis should be a proper tool to compare different innovative solution also in dwelling houses.

Acknowledgements

The authors would like to thank Mrs. and Mr. Lipiński, Office Lipinscy Houses, for making the data of the designed buildings used in the paper available. The comprehensive data on the Lipinscy passive house, built in Smolec near Wrocław, can be found in the catalog of the Bureau Lipinscy Domy.

Translated by Authors

Accepted for print 2.02.2012

References

Collective work. 2003: Guidebook to IEA ECBCS Annex 37: Low Exergy Systems for Heating and Cooling of Buildings, Finland

LIPIŃSCY DOMY. http:// lipinscy.pl, the internet site of the Lipińscy domy Office, 2010.

Sakulpipatsin, P. 2008. Exergy efficient building design, PhD Thesis, TU Delft.

Schmidt D. 2004. *Design of low exergy building – Method and Pre-design Tool.* International Journal of Low energy and Sustainable building, Bd. 3, pp. 1–47.

Szargut J., Petela R. 1965. *Egzergia* [Exergy]. Scientific Technical Publisher (Wydawnictwo Naukowo Techniczne) Warszawa.

Szargut J. 2005. *Exergy method: Technical and Ecological Application.* WITpress.

Torio H., Schmidt D. 2011. ECBCS Annex 49, Low Exergy Systems for High-Performance Buildings and Communities, Final Report.

Wall G. 1999. *Exergetics*, Molndal.

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

GNSS SATELLITE LEVELLING USING THE ASG-EUPOS SYSTEM SERVICES

Karol Dawidowicz

Institute of Geodesy
University of Warmia and Mazury in Olsztyn

Key words: ASG-EUPOS, EUPOS, GBAS, GNSS satellite levelling, geoid.

Abstract

GNSS observations from a network of permanent stations are a complex system offering not only post-processing, but also corrections sent in real time and the creation of virtual observations. In Poland, such a system has been in operation since June 2008: the Polish Active Geodetic Network ASG-EUPOS. For users three services are provided for real-time corrections, and two services are offered for post-processing. In this paper, methods of normal height determination from static GPS measurements were analysed in the context of the technical capabilities of the ASG-EUPOS along with recommendations for such measurements. Particular attention is paid to the possibility of using to such calculations the Virtual Reference Stations (VRS). Studies have shown that height determination using VRS may reduce the length of observation sessions and improve accuracy compared to the results obtained from the NAWGEO or POZGEO services. In addition, because of the short vectors between the virtual station and measured points, accuracy is not dependent on the type of used receiver (L1 or L1/L2).

NIWELACJA SATELITARNA GNSS Z WYKORZYSTANIEM SERWISÓW SYSTEMU ASG-EUPOS

Karol Dawidowicz

Instytut Geodezji
Uniwersytet Warminsko-Mazurski w Olsztynie

Słowa kluczowe: ASG-EUPOS, EUPOS, GBAS, niwelacja satelitarna GNSS, geoida.

Abstrakt

Obserwacje GNSS realizowane na sieciach stacji permanentnych są obecnie złożonymi systemami oferującymi, oprócz postprocessingu, również korekty przesyłane w czasie rzeczywistym oraz tworzenie obserwacji wirtualnych. W Polsce systemem takim jest uruchomiona w czerwcu 2008 roku polska aktywna sieć geodezyjna ASG EUPOS. Dla użytkowników przeznaczono trzy serwisy udostępniania poprawek w czasie rzeczywistym oraz dwa serwisy dla postprocessingu. W pracy przeanalizowano sposoby wyznaczenia wysokości z pomiarów statycznych GPS w kontekście możliwości technicznych systemu ASG-EUPOS oraz niektórych zaleceń do takich pomiarów. Szczególną uwagę zwrócono na możliwość wykorzystania do takich wyznaczeń obserwacji z Wirtualnych Stacji Referencyjnych (VRS – Virtual Reference Station). Przeprowadzone analizy wykazały, że procedura wyznaczenia wysokości punktów z wykorzystaniem VRS może pozwolić na znaczne skrócenie długości sesji obserwacyjnej oraz poprawę dokładności w stosunku do wyników uzyskanych z serwisu NAWGEO czy POZGEO. Ze względu na krótkie wektory między stacją wirtualną a wyznaczanymi punktami dokładność ta nie jest uwarunkowana wykorzystanym w czasie pomiaru typem odbiornika (L1 bądź L1/L2).

Introduction

The EUPOS project (EUropean Position Determination System) was launched in 2002 in Berlin. Its purpose was to create a homogenous ground-based GNSS support system in Central and Eastern Europe. In Poland, ASG-EUPOS (ASG – Aktywna Sieć Geodezyjna) launched in June 2008 (BOSY et al. 2007, 2008). The ASG-EUPOS network assumed the role of a geodetic reference system in Poland. The connection of the ASG-EUPOS stations and the EUREF Permanent Network (EPN) stations (which are located Poland) has allowed the implementation, monitoring and control of the ETRS89 system in the territory of Poland (FIGURSKI et al. 2009).

The ASG-EUPOS is a multi-functional satellite positioning system. Its structure is divided into three basic segments: reference stations, management and user segments. These segments work together to provide a system for precise positioning in real-time and post-processing applications. The network of reference stations (reference segment) currently (as of November 2011) consists of 99 Polish (81 with a GPS module and 18 with a GPS/GLONASS module) and 22 foreign stations (www.asgeupos.pl). The mean distance between reference stations is below 70 km. The stations are regularly distrib-

Table 1

ASG-EUPOS services

Service group	Service name	Survey method	Data access	Estimated precision	Minimum hardware requirements
Real-time services	NAWGEO	kinematic RTK	GSM/ Internet	0.03 m (horiz) 0.05 m (vert.)	L1/L2 GNSS RTK receiver, communication module
	KODGIS	kinematic DGPS		0.2–0.5 m	L1 DGNSS receiver, communication module
	NAWGIS			1.0–3.0 m	
Postprocessing services	POZGEO	static	Internet	0.01–0.10 m	L1 GNSS receiver
	POZGEO D	static/ kinematic			

uted, creating a homogenous network which covers all of Poland. The ASG-EUPOS services enable transfer of reference frame into real applications in the field. Table 1 shows the real-time and post-processing services available in the ASG-EUPOS system (www.asgeupos.pl).

GNSS technology is currently widely used for many kinds of geodetic surveys, including height determination. Relative GNSS positioning encourages users to compute orthometric (normal) height differences, $\Delta H = H_2 - H_1$,by using the well-known relation (Fig. 1):

$$\Delta H = \Delta h - \Delta N \tag{1}$$

where:
$\Delta h = h_2 - h_1$ – the difference in ellipsoidal heights,
$\Delta N = N_2 - N_1$, – the difference in geoid heights.

The accuracy of the calculated ΔH depends on the accuracy of Δh and ΔN. Although it is possible to reach millimetre horizontal relative accuracy levels over tens, or even hundreds of kilometres, vertical GNSS accuracy is not so easily obtained. The baseline vertical component is more sensitive to many disturbing factors, for example: antenna phase centre variations or tropospheric refraction (DAWIDOWICZ 2010, DODSONA et al. 1996, TREGONING, HERRING 2006, WIELGOSZ et al. 2011).

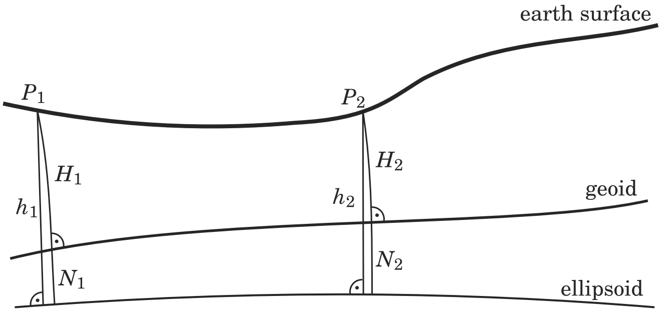


Fig. 1. The concept of GNSS satellite levelling

Source: own work.

The second factor determining the accuracy of GNSS levelling is a geoid (quasigeoid) model. The height system in Poland creates geopotencial numbers divided by the average value of the normal acceleration of gravity along the normal line between the GRS80 ellipsoid and telluroid (further reffered to as "normal heights") referenced to the average level of the Baltic Sea, set for a tide-gauge in Kronstadt near St. Petersburg (Russian Federation) (Rozporządzenie Rady Ministrów... 2000, Projekt nowelizacji RRM... 2007). Normal heights by definition, are related with a quasigeoid. Because in most areas of Poland the spacing between geoid and quasigeoid surfaces is less than 1 cm, sometimes in the literature both surfaces are identified (Banasik 1999, Pażus et al. 2002).

There are a number of categories of techniques for the computation of geoid undulation. Currently the genaral strategy for computation of geoid undulation is composed of the combination of three effects: global, regional and local, which are represented by the geopotential model, mean free-air gravity anomalies and topography respectively (Czarnecki 1994, Hofmann-Wellenhof, Moritz 2005, Torge 1991).

Precise modelling of global and regional geoids has become one of the major tasks of numerous research groups of surveying and mapping agencies. In order to provide determination of normal heights using satellite measurements techniques, the Head Office of Geodesy and Cartography in Poland in 1999 began intensive work on creating a suitable quasigeoid model. The published result of this has been the creation of two quasigeoid models. The first model, called “Geoida niwelacyjna 2000”, is a purely geometric satellite-levelling quasigeoid model based on the hights of the EUREF-POL, POLREF, EUVN, WSSG and Tatry network points. This model was included in the “TRANS-POL” software, which is enclosed in the G1-10 Technical Guidelines. The second published version of quasigeoid was, approved in 2001 by the Surveyor General of Poland for use in geodetic practice, model called “Geoida niwelacyjna 2001”. This model is the result of fitting the “QUASI97B” gravimetric quasigeoid model into the “QGEOID-PL’01” satellite-levelling quasigeoid model based on 752 points, of which 62 belong to the EUVN network, 11 to the EUREF-POL network, 330 to the POLREF network, 23 to the Tatry network and 326 to the WSSG network. A discrete model in the form of quasigeoid heights in 1’ × 1’ grid nodes was determined using the spline function of the third’ degree. Together with the bilinear interpolation formula of quasigeoid heights, it was used in the “GEOIDA” software attached to the Technical Instruction G-2. ASG-EUPOS system is used ‚QGEOID-PG̀ model (Instrukcja Techniczna G-2 2001, Pażus et al. 2002, Wytyczne Techniczne G1-10 2001).

Access to raw gravity data, the development of high-resolution digital terrain models and densification of precise GNSS-levelling heights have simulated extensive research into precise quasigeoid modelling in Poland. Since 2002 a team of reserchers under the leadership of the Institute of Geodesy and Cartography in Warsaw have conducted advanced research into modeling a centimetre quasigeoid in Poland (Kryński, Łyszkowicz 2006a,b).

In the meantime, the release of the EGM2008 (Earth Gravitational Model 2008 EGM2008) by the National Geospatial-Intelligence Agency (NGA) EGM Development Team was a milstone step in precise gravitational and geoidal modelling on a global scale (ŁYSZKOWICZ 2009).

Centimeter-level positioning in an array of reference stations spaced 30 to 100 kilometers apart can be achieved using virtual reference station (VRS) observations. Precise correction models for dispersive (ionospheric) and non-dispersive (tropospheric and orbit) distance-dependent biases are obtained from the real reference data and used in the calculation of the virtual observations (WANNINGER 1997). The process of creating VRS observations based on real observations, consists of several stages (WANNINGER 1997, 1999). The quality of VRS observations fundamentally depends on two aspects: firstly on station-dependent biases in the reference stations observations (mostly caused by a carrier phase multipath) and secondly on possible ionospheric and tropospheric disturbances. Small-scale and medium-scale ionospheric and tropospheric refraction may not be completely represented by the correction models of distance-dependent biases. The remaining errors affect ambiguity resolution and positioning accuracy of the baseline between VRS and the measured point. The size of these biases (and thus the quality of the VRS observations) is estimated together with the correction model parameters (WANNINGER 1999).

The VRS concept was initially dedicated to RTK (Real Time Kinematic) measurements and is now more often used in a static approach (BAKUŁA 2006, SIEJKA 2009).

Performing geodetic surveying is regulated by a number of legal acts and technical standards. At the moment there are no definitive instructions for measurements using the ASG-EUPOS. In the project phase are technical guidelines G-1.12. At the beginning of 2011, the Surveyor General of Poland published the "GNSS satellite measurements based on reference stations of ASG-EUPOS" technical recommendations, which cannot be treated as a valid technical standard for GNSS measurements. At present, the use of ASG-EUPOS services is possible, as long as it meet accuracy requirements. In such a situation, it is necessary to conduct detailed studies related to the determination of position using ASG-EUPOS.

In this paper, methods of normal heights determination from GNSSS measurements were analysed in the context of the technical capabilities of the ASG-EUPOS along with recommendations for such measurements. Particular attention was paid to the possibility of using VRS observations for such determination.

Research area

Four points situated in Kortowo (UWM Olsztyn) were selected for test measurements (Fig. 2). This location resulted in the nearest CORS stations of the ASG-EUPOS system being between 3 and 20 km distant (Fig. 3).

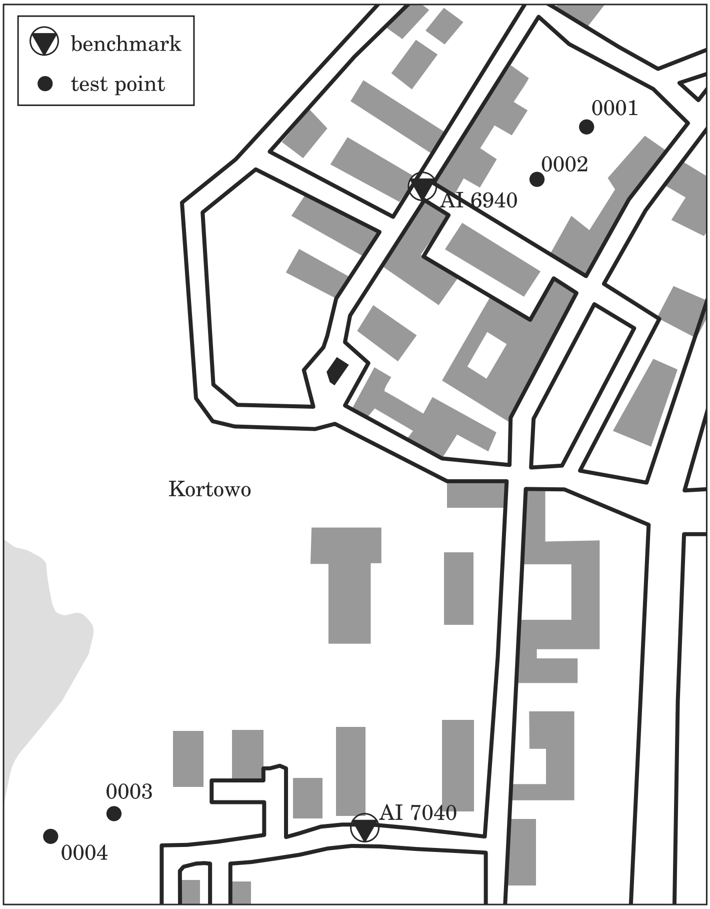


Fig. 2. The location of existing levelling benchmark and test points
Source: own work.

One hundred independent RTK measurements were performed on each test points with the NAWGEO service. Between the measurements, random breaks were made from 5 to 30 seconds. Additionally a static session was carried out on the test points . The following GNSS parameters were assumed for that session: sampling interval 5s, minimum satellite; elevation 10°, time of measurement 4 hours.

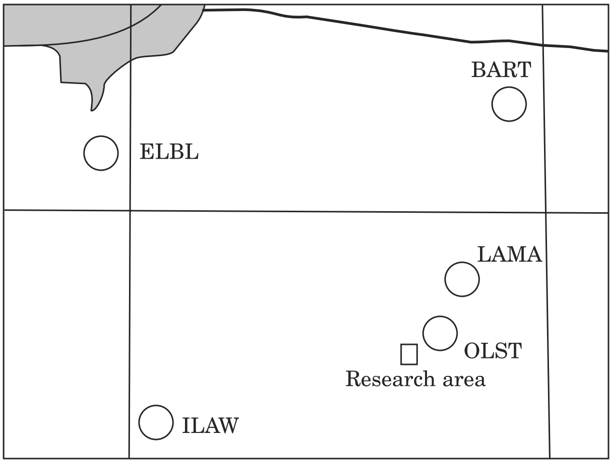


Fig. 3. The location of the research area
Source: map: http://www.asgeupos.pl

For the test points normal heights were determined by precise leveling (Tab. 2). Precise levelling was assigned to two 2nd order benchmarks of the national levelling network (AI 6940, AI 7040). Table 2 also includes the separations between quasigeoid and ellipsoid on measured points calculated with three quasigeoid models (among others the "QGEOID-PG" model which is used in the ASG-EUPOS system).

Table 2

Normal height and quasigeoid-to-ellipsoid separation on measured points

Measurement point number	Normal height [m]	Quasigeoid to ellipsoid separation from "Geoida niwelacyjna 2000" model [m]	Quasigeoid to ellipsoid separation from "Geoida niwelacyjna 2001" model [m]	Quasigoid to ellipsoid separation from "QGEOID-PG" model [m]
0001	117.234	29.816	29.815	29.805
0002	115.318	29.815	29.814	29.803
0003	104.995	29.831	29.829	29.818
0004	105.974	29.831	29.830	29.819

On all test points about a 1 cm difference between quasigeoid-to-ellipsoid separations calculated from "QGEOID-PG" model and both other models is visible. The impact of the quasigeoid model used for the normal height determination has been analysed, among others, by HADAŚ and BOSY (2009).

Analysis of results

Results of the RTK-NAWGEO measurements (normal height differences between heights obtained from precise levelling and heights obtained from satellite leveling), are presented in figure 4 (the "QGEOID-PG" model was used).

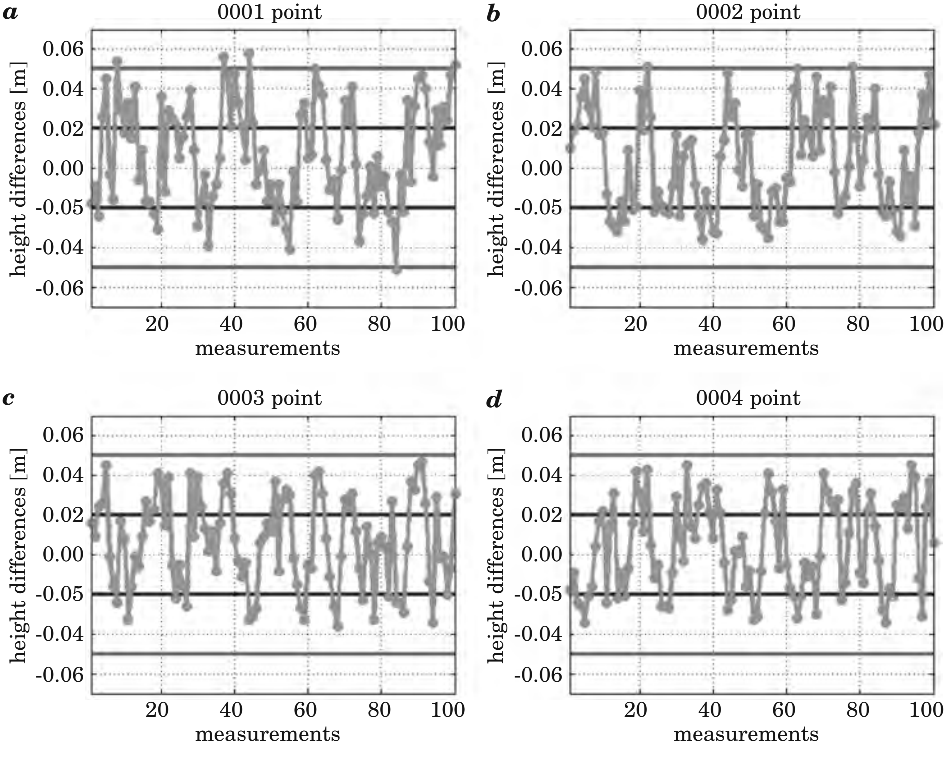


Fig. 4. Normal height differences between heights obtained from precise levelling and heights obtained from RTK-NAWGEO measurements

Figure 4 shows, that for measured points, on average, within ±2 cm was nearly half the height differences (43%), while in the range ±5 cm (accuracy of the RTK height determination declared by the ASG-EUPOS system): 98% height differences. It should be noted that the largest difference was observed for the points of nearby tall trees and buildings (0001 and 0002 points). The standard deviation, being a measure of the precision of the data, was 1.4 cm.

Observations from static measurements carried out on test points have been processed in several variants. Because cost-effectiveness is a requirement for most geodetic projects several analyses were conducted into how the accuracy depends on the baseline length and on the duration of the observing session (ECKLE et al. 2001, PSIMOULIS et al. 2004). For this reason, four-hour session was divided into 4 one-hour sessions and 8 half-hour sessions. Next, the four and one-hour sessions were sent to the POZGEO service. Additionally, all measurement sessions were processed with Topcon Tools v7.3 software (POZGEO-D service) in three variants: with LAMA, with OLST and with VRS stations as reference stations. VRS station were created with the POZGEO-D service in the immediate vicinity of the measured points.

In Topcon Tools software, selection of processing frequency is automatic and for baselines up to 30 km, it appears as follows: 0–10 km processing on L1 frequency, 10–30 km processing in ionosphere-free combinations. To process the IGS final orbit and IGS the absolute (converted from relative) antenna phase center offset model was chosen. The results for one and half hour sessions (ellipsoidal heights and their RMS errors) are presented in figures 5–8.

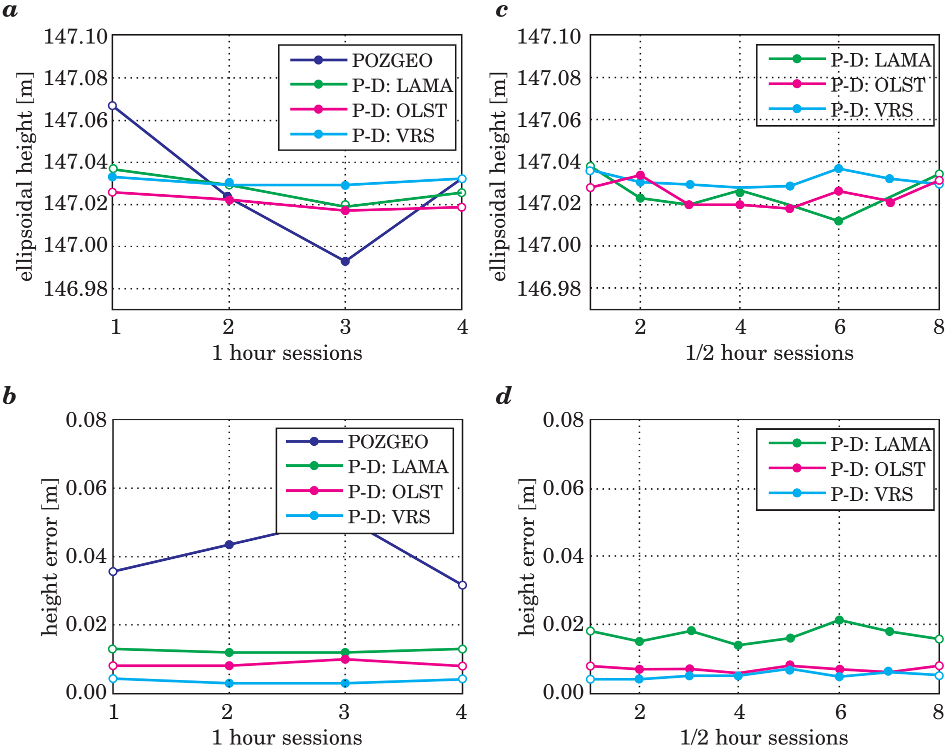


Fig. 5. Ellipsoidal height and height error for 0001 point: *a, b* – one hour sessions, *c, d* – half hour sessions

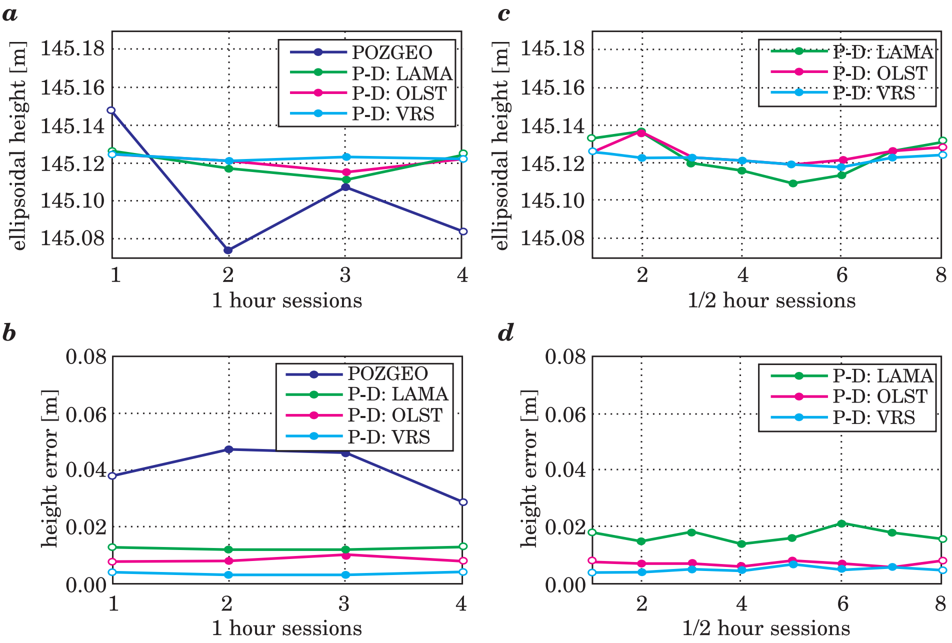


Fig. 6. Ellipsoidal height and height error for 0002 point: *a, b* – one hour sessions, *c, d* – half hour sessions

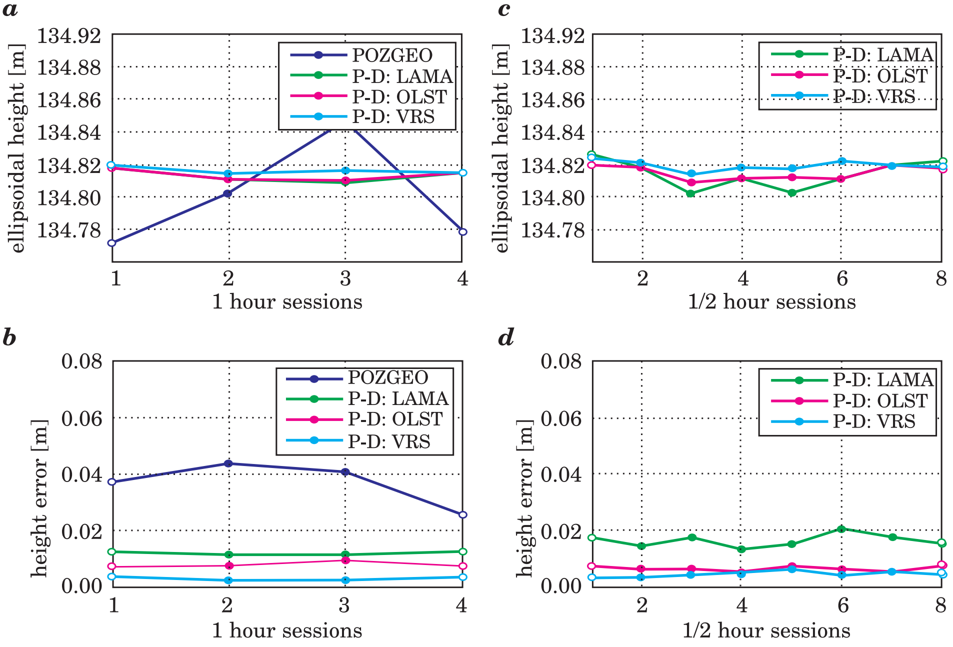


Fig. 7. Ellipsoidal height and height error for 0001 point: *a, b* – one hour sessions, *c, d* – half hour sessions

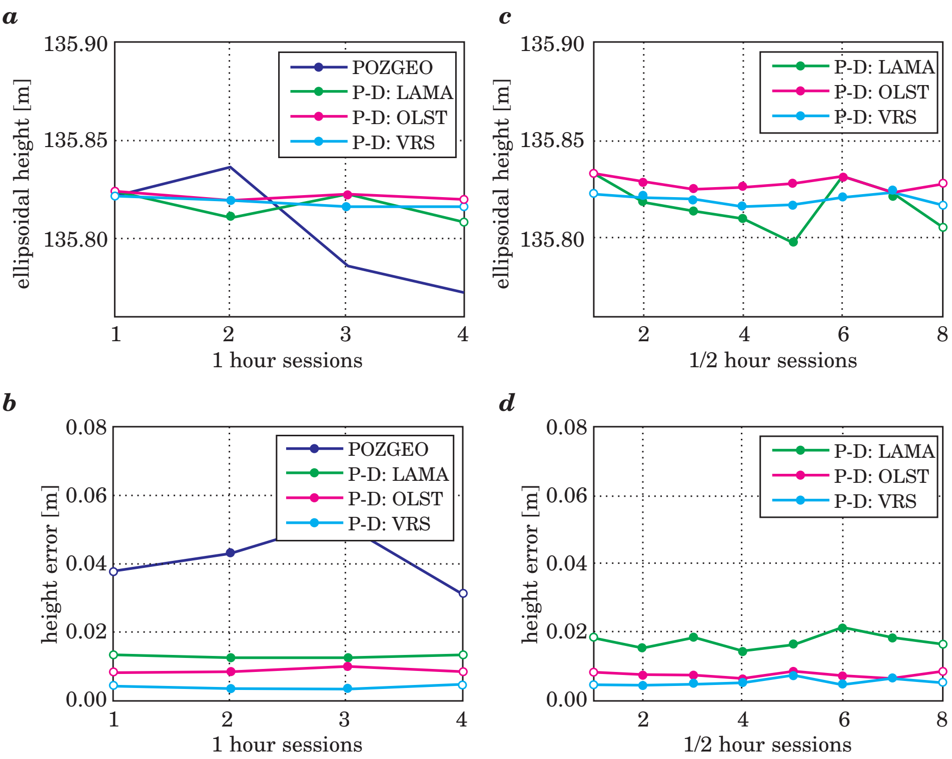


Fig. 8. Ellipsoidal height and height error for 0002 point: *a, b* – one hour sessions, *c, d* – half hour sessions

Anlysing the results on figures 5–8, it is visible that the ellipsoidal heights obtained from the POZGEO service are characterized by the lowest stability and the highest value of RMS error (one hour session variants – min. 720 epoch needed). It is clear that there is a tendency to increase stability and reduce the RMS error by shortening the distance between measured points and reference station. The best results were obtained for the variants using the OLST or VRS stations as reference points. It should be noted that this solution was obtained for L1 processing.

For all processing variants the normal heights were calculated. The "QGEOID-PG" model was used to calculate the distances between the quasigeoid and ellipsoid. The normal height differences between heights obtained from precise levelling and the heights obtained from satellite levelling are presented in figure 9.

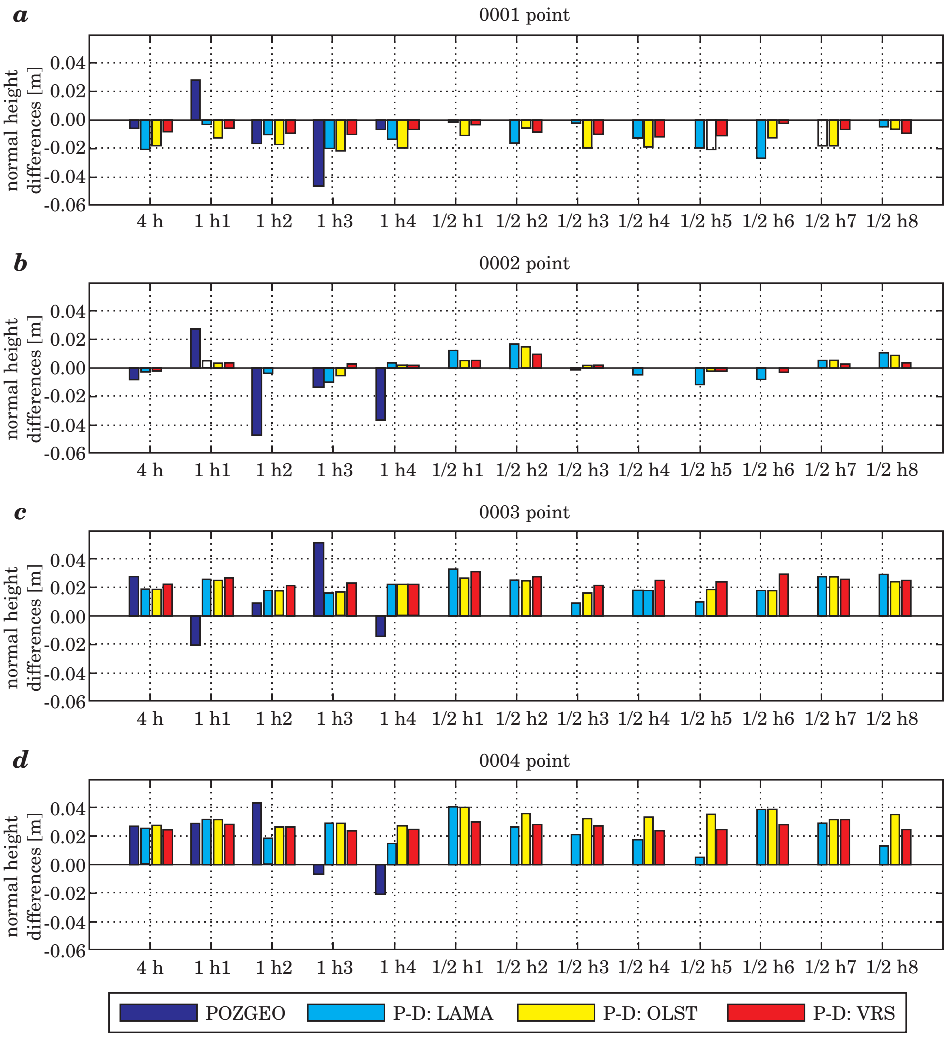


Fig. 9. Normal height differences between heights obtained from precise levelling and heights obtained from satellite levellin

Height differences calculated from the results from one-hour sessions were within ±2 cm range for 44% of the POZGEO solution, for 75% of the POZGEO-D LAMA (P-D LAMA) solution, for 56% of POZGEO-D OLST (P-D OLST) solutions and for 50% of POZGEO-D VRS (P-D VRS) solutions. With ±3 cm range were 69% of the POZGEO solution, 94% of the POZGEO-D LAMA solution, 94% of POZGEO-D OLST solutions and 100% of POZGEO-D VRS solutions.

Height differences calculated from the results of half-hour sessions were within ± 2 cm range for 68% of the POZGEO-D LAMA solution, for 59% of POZGEO-D OLST solutions and for 50% of POZGEO-D VRS solutions. With ±3 cm range were 90% of the POZGEO-D LAMA solution, 75% of POZGEO-D OLST solutions and 90% of POZGEO-D VRS solutions.

Conclusions

This paper presents an analysis of normal height determination using the services of the ASG-EUPOS system (NAWGEO, POZGEO, POZGEO-D).

In the case of RTK satellite levelling, the low accuracy of ellipsoidal height determination (dispersion of the range of 10 cm) has a significant impact on the value of normal height calculation. A one-hour static session also proved to be too short for accurate determination of heights in the POZGEO service (ellipsoidal height dispersion in the range of 7 cm). Neither method provides determination of normal heights with ±2–3 cm accuracy.

A better solution at the moment is either extending the measuring session for processing in the POZGEO service or own observation post-processing (POZGEO-D service). A half-hour session was sufficient to determine ellipsoidal heights with ±1 cm accuracy (variants with OLST or VRS as reference station). Additionally, because of the short vectors between the measured points and the reference station, the accuracy of the calculations did not depend on the type of receivers used for the measurement (L1 or L1/L2). For those processing variants, the accuracy of quasigeoid model has a significant impact on the value of normal height calculation. It is hoped that this accuracy will be improved with the newest quasigeoid model for Poland.

Translated by Joanna Jensen

Accepted for print 12.01.2012

References

Bakuła M. 2006. *Performance of Static Positioning for Medium Distances Based on Data from Virtual Reference Station and ASG-PL Network*. Artificial Satellites, 41(1): 33–42.

Bosy J., Graszka W., Leonczyk M. 2007. *ASG-EUPOS – a Multifunctional Precise Satellite Positioning System in Poland*. International Journal on Marine Navigation and Safety of Sea transportation. December 2007, 7(4): 371–374.

Bosy J., Oruba A., Graszka W., Leończyk M., Ryczywolski M. 2008. A*SG-EUPOS densification of EUREF Permanent Network on the territory of Poland*. Reports on Geodesy, 2(85): 105–112.

Banasik P. 1999. *Wyznaczenie przebiegu quasi-geoidy w rejonie Krakowa na podstawie pomiarów niwelacyjnych i GPS*. Rozprawa doktorska. Kraków.

Czarnecki K.1994. *Geodezja współczesna w zarysie*. Wydawnictwo Wiedza i Życie, Warszawa.

Dawidowicz K. 2010. *Antenna phase centre variations corrections in processing of GPS observations with use of commercial software*. Technical Sciences, 13: 120–132.

DODSON A.H., SHARDLOW P.J., HUBBARD L.C.M., ELGERED G., JARLEMARK P.O.J. 1996. *Wet tropospheric effects on precise relative GPS height determination*. Journal of Geodesy, 70: 188–202.

ECKLE M.C., SNAY R.A., SOLER T., CLINE M.W., MADER G.L. 2001. *Accuracy of GPS-derived relative positions as a function of interstation distance and observing-session duration*. Journal of Geodesy, 75: 633–640.

FIGURSKI M., KAMIŃSKI P., KROSZCZYŃSKI K., SZAFRANEK K. 2009. *ASG-EUPOS monitoring with referemce to EPN*. Artificial satellites, 44(3): 85–04.

HADAŚ J., BOSY J. 2009. *Niwelacja satelitarna GNSS z wykorzystaniem serwisu NAWGEO system ASG-EUPOS*. Acta Sci. Pol., Geodesia et Descripto Terrarum, 8(2): 53–66.

HOFMANN-WELLENHOF B., LICHTENEGGER H., WASLE E. 2008. *GNSS – Global Navigation Satellite Systems*. Springer-Verlag Wien, Austria.

HOFMANN-WELLENHOF B., MORITZ H. 2005. *Physical Geodesy*. Springen, Wien-NewYork.

Instrukcja Techniczna G-2. 2001. *Szczegółowa pozioma i wysokościowa osnowa geodezyjna i przeliczanie współrzędnych miedzy układami*. Główny Urząd Geodezji i Kartografii, Warszawa.

KRYŃSKI J., ŁYSZKOWICZ A. 2006a. *Regional quasigeoid determination in the area of Poland*. 5th FIG regional Conference for Africa, Accra, Ghana, 8–11 March.

KRYŃSKI J., ŁYSZKOWICZ A. 2006b. *Centimetre quasigeoid modelling in Poland using heterogenous data*. IAG Proceedings from 1st International Symposium of the International Gravity Field Service (IGFS), 28 August – 1 September, Istanbul, Turkey.

ŁYSZKOWICZ A. 2009. *Oszacowanie dokładności quasi-geoidy z modelu EGM08 na obszarze Polski.* Acta Scientarum Polonorum, Geodesia et Descripto Terrarum, 8(4): 39–47.

PAZUS R., OSADA E., OLEJNIK S. 2002. *Geoida niwelacyjna 2001*. Magazyn Geoinformacyjny GEODETA, 5(84).

Projekt nowelizacji rozporządzenia Rady Ministrów z dnia 8 sierpnia 2000 r. w sprawie państwowego systemu odniesień przestrzennych (Dz.U. nr 70, poz. 821).

PSIMOULIS P.A., KONTOGIANNI V.A., NICKTITOPOULOU A., PYTHAROULI S.I., TRIANTAFYLLIDIS P., STRIOS S.C. 2004. *Estimating the Optimum Duration of GPS Static Observations for Short Baseline Length Determination in Greece*. Proceedings of the FIG Working Week, Athens, Greece, 22–27 May.

Rozporządzenie Rady Ministrów z dnia 8 sierpnia 2000 r. w sprawie państwowego systemu odniesień przestrzennych (Dz.U. nr 70, poz. 821).

SIEJKA Z. 2009. *Wykorzystanie pomiarów GNSS do wyznaczania współrzędnych podstawowej osnowy realizacyjnej na terenach oddziaływań górniczych*. Archiwum fotogrametrii, Kartografii i Teledetekcji, 19: 387–396.

THREGONING P., HERRING T.A. 2006. *Impact of a priori zenith hydrostatic delay errors on GPS estimates of stadion heights and zenith total delays.* Geophysical Research Letters, 33: L23303.

TORGE W. 1991. *Geodesy.* 2nd Edition, Walter de Gruyter Berlin-New York.

Topcon Positioning System. 2006. Topcon Tools User's Guide. Topcon Positioning Systems Inc, May.

WANNINGER L. 1997. *Real-Time Differential GPS-Error Modelling in regional Reference Stations Networks*. Proc. IAG Symp 118, Rio de Janerio, pp. 86–92.

WIELGOSZ P., PAZIEWSKI J., BARYŁA R. 2011. *On constraining tropospheric delays in the processing of local GPS networks with BERNESE software.* Survey Review, 43(323): 472–483.

Wytyczne techniczne G-1.12. 2008. *Pomiary satelitarne oparte na systemie precyzyjnego pozycjonowania ASG-EUPOS.* http://geogis.com.pl/download/wytyczne_g_1_12_21_04_2008_1.pdf.

Wytyczne techniczne G1-10. 2001. *Formuły odwzorowawcze i parametry układów współrzędnych*. Główny Urząd Geodezji i Kartografii, Warszawa.

Zalecenia techniczne. 2011. *Pomiary satelitarne GNSS oparte na systemie stacji referencyjnych ASG-EUPOS.* http://www.asgeupos.pl/webpg/graph/standards/Zalecenia_ASG_EUPOS_20110207.pdf.

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

GEOID IN THE AREA OF POLAND IN THE AUTHOR'S INVESTIGATIONS

Adam Łyszkowicz

Chair of Land Surveying
University of Warmia and Mazury in Olsztyn

Key words: geopotential models, geoid, quasigeoid, deflections of the vertical.

Abstract

The present paper describes the results of the author's work related to the geoid determination in the area of Poland. Beginning from the geoid model *geoid92* worked out in 1992, various geoid models calculated from various data sets, using various methods, are presented. Additionally, the evaluation of the accuracy of the determined geoid models is given. Next, the necessity of the fitting of computed geoid models to the national vertical reference system and the evaluation of the accuracy of the functionals N, Δg, ξ, η of the gravity field calculated from the various geopotential models are presented.

It results from the presented investigations, that at present the accuracy of gravimetric geoid/quasigeoid models is ± 1.4 cm and accuracy of geoid computed from geopotential model EGM08 is ±2.4 cm. It is stated additionally that accuracy of gravity anomalies computed from the EGM08 model is ± 25 μm s^{-2} , while accuracy of the deflections of the vertical ξ and η is from ±0.6'' to ± 0.7''.

BADANIE PRZEBIEGU GEOIDY NA OBSZARZE POLSKI

Adam Łyszkowicz

Katedra Geodezji Szczegółowej
Uniwersytet Warmińsko-Mazurski w Olsztynie

Słowa kluczowe: modele geopotencjału, geoida/quasi-geoida, składowe odchyleń pionu.

Abstrakt

W pracy przedstawiono wyniki badań dotyczących przebiegu geoidy na obszarze Polski. Poczynając od modelu *geoid92,* opracowanego w 1992 roku, w artykule przestawiono kolejne modele geoidy liczone z różnych danych i różnymi metodami wraz z ich charakterystyką dokładnościową. Uzasadniono konieczność dopasowania wyliczonych modeli geoidy do krajowego układu wysokościowego oraz podano dokładności charakterystyk pola siły ciężkości N, Δg, ξ, η obliczanych z kolejnych modeli geopotencjału.

Z wyników badań wynika, że obecnie dokładność grawimetrycznych modeli geoidy/quasi-geoidy wynosi ±1,4 cm, a dokładność geoidy wyliczonej z modelu geopotencjału EGM08 ±2,4 cm. Dodatkowo stwierdzono, że dokładność anomalii grawimetrycznych wyliczonych z modelu EGM08 wynosi ±25 μm s^{-2}, a dokładność składowych odchyleń pionu ξ i η od ±0.6'' do ±0.7''.

Introduction

Geoid has the basic role in geodesy, oceanography and geophysics. It serves in geodesy and oceanography, as the reference surface describing the topography of continents and oceanic surfaces. One uses geoid in geophysics, as the representation of the gravity field which reflects the expansion of earth mass situated in the depth of the Earth.

The paper is based on report presented at the conference dedicated to the fifth anniversary of Institute of Geodesy and Geoinformatics in Wroclaw. The main aim of this paper is to show, how in author's investigations, methods and data in geoid/quasigeoid models computation in the area of Poland has changed and how the accuracy of successive models have increased. Beside geoid determination it presents computation of gravimetric deflections of the vertical e.g. (ŁYSZKOWICZ 1996a) and recapitulations of these investigations are also in this paper. More and better geopotential models create a new possibility in getting the gravity anomalies and the deflections of the vertical without necessity of laborious field measurements. In the present paper the newest author's investigations in this matter are presented.

The paper consists of nine sections. After the introduction, in the second section the accuracy evaluation of geoid undulations, deflections of the vertical and gravimetric anomalies are presented. In the next section, briefly an overview of the basis of geoid calculation from gravimetric data and then geoid models computed by the author in the last 20 years are presented. Section four contains information relating to the determination of geoid from the deflections of the vertical by the collocation method. Section five deals with the necessity of fitting computed geoid models to the vertical reference system, and then various ways of such fitting are considered. Then it is shown how the accuracy of geoid computed from various geopotential models increased in the last 20 years. Last section contains recapitulation and conclusions.

Evaluation of geoid accuracy

The estimation of geoid/quasigeoid accuracy was for many years unusually difficult. At the moment when appeared the possibility of the establishment of geodetic networks by the satellite GPS method appeared, the estimation of geoid accuracy simplified considerably and it looks as follows.

The relationship between geoid undulation N, ellipsoidal height h and orthometric height H is.

$$h = H + N \tag{1}$$

From this formula it follows, that if the points of a satellite network at which ellipsoidal heights h were determined from satellite observation are tied to the national levelling network, then the geoid/quasigeoid heights at these points can be computed from relationship.

$$N_i^{\text{gps/levelling}} = h_i - H_i \tag{2}$$

From the available (e.g. gravimetric) geoid/quasigeoid models, the distance between geoid and ellipsoid N_i^{grav} can be computed at the points of a satellite network and then the differences can be created

$$\delta N_i = N_i^{\text{gps/levelling}} - N_i^{\text{grav}} \tag{3}$$

from which mean value $\hat{x}$ and the empirical standard deviation $\hat{\sigma}_N$ can be computed. So computed empirical standard deviation $\hat{\sigma}_N$ is an estimation of absolute accuracy of geoid models. The method of estimation of the relative accuracy of geoid/quasigeoid models is more complicated and is given e.g. in the paper (Łyszkowicz 2009a).

In a case of estimation of gravity anomalies Δg_{GM} computed from geopotential models, these anomalies are compared with gravity anomalies Δg obtained from terrestrial measurements and then the differences are created as

$$\delta\Delta g_i = \Delta_i^{\text{GM}} - \Delta g_i \tag{4}$$

Empirical standard deviation $\hat{\sigma}_{\Delta g}$ of differences is an estimate of the accuracy of gravimetric anomalies computed from geopotential models.

In similar way is estimated the accuracy the deflections of the vertical computed from geopotential model or gravity data is estimated

$$\begin{aligned} \delta\xi_i &= \xi_i^{\text{GM}} - \xi_i \\ \delta\eta_i &= \eta_i^{\text{GM}} - \eta_i \end{aligned} \tag{5}$$

where ζ_i^{GM}, η_i^{GM} are computed from a geopotential model or gravity data and ζ_i, η_i are astrogeodetic deflections of the vertical. Empirical standard deviations $\hat{\sigma}_\xi$, $\hat{\sigma}_\eta$ of differences are estimate of the accuracy of deflections.

In the present work to estimate the accuracy of geoid/quasigeoid models the POLREF network and control traverse were used. POLREF network consists of 360 points measured in campaigns from July 1994 till May 1995 (Zieliński et al. 1997). Traverse is about 868 km length, running through the whole area of Poland, consisting from 190 points, measured in five campaigns in the years 2003–2004 (KRYŃSKI et al. 2005). The points of the POLREF network and traverse were tied up to the precise levelling network whose heights are expressed in Kronsztadt86 system.

The gravimetric data used for the evaluation of accuracy of anomalies computed from successive geopotential models and also to calculate the successive versions of gravimetric geoid/quasigeoid models contains data from the following area: Poland, Czech Republic, Slovakia, Hungary, Romania, Ukraine, Germany and Denmark and they are describes in detail in (ŁYSZKOWICZ 1994).

To estimate the accuracy of deflections of the vertical determined from geopotential models and to compute astrogeodetic geoid by the collocation method, 171 astrogeodetic deflections of the vertical measured during years 1967–1981 were used (*Katalog* 1981). These deflections were then improved (ROGOWSKI et al. 2003) because of the new star catalogue, movements of the pole, time system and horizontal reference system and these improved deflections were used in our computations.

Geoid computation from Stokes integral and by the collocation method

The general adopted strategy of the geoid/quasigeoid calculation consists of combination of three effects: global, regional and local which are represented suitably through geopotential model GM, mean residual Faye anomalies and the topography

$$N = N_{\mathrm{GM}} + N_{\Delta g_{\mathrm{res}}} + N_H \tag{6}$$

Term N_{GM} represents the contribution of spherical harmonic coefficients, while the term $N_{\Delta g_{\mathrm{res}}}$ represents the contribution of residual Faye anomalies after removing the effects due to the geopotential model

$$\Delta g_{\mathrm{res}} = \Delta g_{\mathrm{FA}} - \Delta g^{\mathrm{GM}} \tag{7}$$

The term N_H in equation (6) is defined, as the indirect effect on geoid and represents the change of the equipotential surface after applying terrain reduction to Δg.

Component $N_{\Delta g_{\mathrm{res}}}$ can be computed from Stokes integral which is evaluated by the fast Fourier transforms (FFT) or by the collocation method. Computed geoid undulations can be when transform into quasigeoid according to the equation

$$\zeta = N - \frac{\Delta g_B}{\gamma} H \tag{8}$$

where Δg_B is Bouguer anomaly, γ is mean normal gravity and H is topographic height.

Gravimetric geoid models

In the present chapter are presented the successive versions of gravimetric geoid/quasigeoid models, which were computed by the author for the area of Poland in the last 20 years. In the description of the models, we pay attention to gravity data, methods of computation and obtained accuracy.

Model geoid92

The first gravimetric geoid model for Central Europe, including Poland, was computed in 1949 (TANNI 1949) and has very low accuracy. The first gravimetric geoid model for the area of Poland was computed at the Department of Planetary Geodesy, Polish Academy of Sciences in 1993. This model was computed by the combination of collocation and the integral method and it is based on the OSU81 geopotential model to degree and order 80 and on 6000 Faye gravity anomalies from the area of Poland. The estimation shows that the accuracy of this model is on the level of ± 26 cm (ŁYSZKOWICZ 1993)[1].

Model geoid94

The second gravimetric geoid model for the area of Poland *geoid94* was computed using Stokes integral which was evaluated by the FFT method and is based on a much bigger sets of gravimetric data (about 8000 additional mean anomalies), topographic data and geopotential model OSU91 to degree and order 360 (ŁYSZKOWICZ, DENKER 1994). The accuracy of this model is estimated[2] as ±14 cm.

Model quasi95

In 1994 within realization of the research project of Committee of Scientific Research (CSR) (ŁYSZKOWICZ 1995) a new set of mean Faye anomalies was used

[1] The evaluation of this model was done in 1993 on 10 points of the EUREF network, what means that estimation of accuracy is not realistic.

[2] New evaluation done in 2010 on the basis of all points of POLREF network.

in the geoid computation for the area of Poland in a grid 1' × 1'. These data together with gravimetric data collected from neighbouring countries and digital terrain model in a grid 1.5' × 3' interpolated from the terrain model in a grid 30' × 30' for Poland, and 5' × 7.5' for the neighbouring areas were used to compute the first gravimetric quasigeoid model named *quasi95* (ŁYSZKOWICZ, FORSBERG 1995). The calculation of this model was conducted with the utilization of fast Fourier transforms (FFT). In these calculations the global geopotential model OSU91 was used to degree and order 360. The accuracy of this model is estimated[3] as ± 8.7 cm.

This model was fitted to the vertical reference system Kronsztadt86 (section 7) and was made accessible to practical use by Head Office of Geodesy and Cartography in 1996 (ŁYSZKOWICZ 1997).

Model quasi97b

Next quasigeoid model named *quasi97b* was computed in 1997 from at least 140 000 point and mean gravity anomalies and a new geopotential model EGM96 to degree and order 360 (LEMOINE at al. 1998) by fast Fourier transforms (FFT) method. This model was realized within the project ordered by Committee of Scientific Research (ŁYSZKOWICZ 1998) and together with suitable software it was delivered to Head Office of Geodesy and Cartography to be used in geodetic practice. The accuracy of this model is estimated as ± 3.5 cm.

In order to use in practice *quasi97b* model, it was fitted to vertical reference system Kronsztadt86 by the collocation method. Accuracy of fitted *quasi97b* model is estimated as ± 1.0 cm (ŁYSZKOWICZ 2000)[4].

Model quasi09a

Computed in 2009 the next version of gravimetric quasigeoid model named *quasi09* was developed on the basis of gravimetric data, which consists of the set of mean anomalies from the area of Poland (KRYŃSKI 2007) and of data from neighbouring countries. Additionally during the calculations topographical data were used, that is DTED (*Digital Terrain Elevation Data*) model and SRTM (*Shuttle Radar Topography Mission*) model. The gravimetric model *quasi09a* was computed with the use of gravity data described above and geopotential model EGM08 to degree and order 2190 (PAVLIS et al. 2008) by the fast Fourier method (ŁYSZKOWICZ 2009b).

[3] New evaluation done in 2010 on the basis of all points of POLREF network.

[4] This evaluation is probably too optimistic.

Accuracy of *quasi09a* model estimated at the points of POLREF network gives a mean error of ± 3.5 cm, while analogous estimation at the points of the traverse gives a mean error ± 2.7 cm.

Model quasi09c

In 2009, for the first time in Poland, a gravimetric quasigeoid model (*quasi09c*) was computed by the collocation method and, for the first time the mean square error of the term $N_{\Delta g_{\mathrm{res}}}$ present in formula (6) was estimated. The error of this component over the area of Poland is from 0.3 cm to 0.4 cm (ŁYSZKOWICZ 2010a).

Estimation of the accuracy of *quasi09c* model on the points of the network POLREF gives the mean error ±3.2 cm, while analogous judgment conducted on the points of the traverse gives the mean error ±1.8 cm (ŁYSZKOWICZ 2010a). Additional data in the form of 171 astrogeodetic deflections of the vertical did not improve noticeably the accuracy of this model.

Geoid "GUGiK 2001"

The next published version of the quasigeoid model approved in 2001 by the Head Office of Geodesy and Cartography to be used in geodetic practice was the model named "*GUGiK 2001*" (PAŻUS 2001). This model came into being as a result of the fitting to the vertical reference system Kronsztad86 the gravimetric quasigeoid model *quasi97b* using spline functions of degree 3 (PAŻUS et al. 2002).

In the present work estimation of the accuracy of this model was done at 140 points of the traverse and we get the mean square error of this model ±1.8 cm. Estimation of accuracy of the deflection of the vertical computed from this model was done on 171 points of the astrogeodetic network and we got a mean error of the components ξ and η as 0.60'' and 0.68'' second of arc respectively.

Graphical illustration of the accuracy of the various gravimetric geoid/quasigeoid models are given in figure 1.

Geoid from deflections of the vertical by Helmert and collocation methods

The aim of this investigation was to use the first time in Poland the collocation method to determine a geoid model from astrogeodetic deflections of the vertical and show that this method is significantly better than the classical Helmert method.

Deflections of the vertical are traditionally used to the determination the geoid in a local or regional scale. The first astrogeodetic geoid model for the area of Poland was determined in 1961 (BOKUN 1961) and the next astrogeodetic geoid model was determined in 2005 (ROGOWSKI et al. 2005). This last model was an improved one due to the removal of systematic errors, in astrogeodetic and astrogravimetric deflections of the vertical.

The latest improved astrogeodetic geoid model is presented in the paper (ŁYSZKOWICZ 2010b). It was computed for the area of Poland, by the use improved data and a better computational algorithm, the collocation method. The theoretical fundamentals of the astronomical levelling and collocation method are given briefly e.g. in paper (ŁYSZKOWICZ 2010b, sec. I). In the same paper an evaluation of the accuracy of the deflections of the vertical and weights estimation is given.

Section II of the paper (ŁYSZKOWICZ 2010c) describes four computed geoid models, namely: two models from the improved astrogeodetic and astrogravimetric data by the Helmert method and models from the same two data sets but computed by the least squares collocation method.

These models were compared with quasigeoid undulation $N^{\text{gps/levelling}}$ at the points of the satellite POLREF network. The results of the comparison show, that geoid model determined by the collocation method is characterize by an accuracy of ±12 cm – ±7 cm and is 5–7 times better than the accuracy of the models computed by the classical method.

Geoid fitting to the vertical reference system

In practice the relationship (1) is not fulfiled and each component in equation (1) is affected by systematic errors that can be written in the form

$$h + \delta h = (H + \delta H) + (N + \delta N) \tag{9}$$

Equation (9) can be written in the form

$$h = H + N + (\delta H + \delta N - \delta h) \text{ or } h = H + N + c \tag{10}$$

where c is correction surface that enables more precise transformation of heights H from spirit levelling into ellipsoidal heights h from satellite GPS observations and vice versa. Correction surface c can be determined from certain parametric models given in table 1 and additionally by the collocation method e.g. (ŁYSZKOWICZ 2000).

Model#1 in table 1 contains one parameter which can be interpreted as a constant shift between the considered surfaces. Model#1 was used to assess the accuracy of geoid models described in the previous sections of this paper.

Model#2 contains two additional parameters which represent mean north-south and east-west inclination e.g. between the gravimetric geoid and geoid from GPS and levelling.

Model#3 is commonly known as four parameter model which geometrically represents three shifts along the x, y z axis and scale change of reference systems i.e. geoid heights with respect to GPS heights or the opposite.

Finally, models #4, #5 and #6 represent height-dependent linear corrector surfaces that constrain the relation between ellipsoidal, normal and quasigeoidal heights in terms of the generalized equation

$$h - (1 + \delta_{S_H})H - (1 \ \delta_{S_H}) = a_0 \tag{11}$$

The above equation takes into consideration the fact that the spatial scale of GPS heights does not necessarily match with the spatial scale induced by quasigeoid undulations of the geopotential model quasigeoid undulations and/or the inherent scale of normal heights obtained from spirit levelling. Moreover, the geopotential model of quasigeoid heights and/or local normal heights are often affected by errors that are correlated, to a certain degree, with the Earth's topography, what is a fact that can additionally justify the use of model #4 or #6 for optimal fitting between $\zeta^{\mathrm{gps/levelling}}$ and ζ.

Table 1

List of various parametric models, KOTSAKIS (2008)

Number of the model	Type
1	$a_0 + (h_i - H_i - N_i) = v_i$
2	$a_0 + a_1 (\varphi_i - \varphi_0) + a_2 (\lambda_i - \lambda_0) \cos\varphi_i + (h_i - H_i - N_i) = v_i$
3	$a_0 + a_1 \cos\varphi_i \cos\lambda_i + a_2 \cos\varphi_i \sin\lambda_i + a_3 \sin\varphi_i + (h_i - H_i - N_i) = v_i$
4	$a_0 + \delta_{S_H} H_i + (h_i - H_i - N_i) = v_i$
5	$a_0 + \delta_{S_H} H_i + (h_i - H_i - N_i) = v_i$
6	$a_0 + \delta_{S_H} H_i + \delta_{S_N} + (h_i - H_i - N_i) = v_i$

In the paper (ŁYSZKOWICZ 2009a) the accuracy of fitting of the quasigeoid undulation computed from the EGM08 model to the vertical reference system Kronsztadt86 with the use of various parametric models (Table 1) was tested. From these investigations it can be concluded that the model #2 and #3 which are commonly used in practice e.g. (FORSBERG 1998) give somewhat better results than the remaining models. The best results are given by the parametric model #3 which makes possible to fit the quasigeoid from EGM08 to the vertical reference system in Poland with the accuracy ±2.4 cm (ŁYSZKOWICZ 2009a) and therefore it will be used in our further calculations.

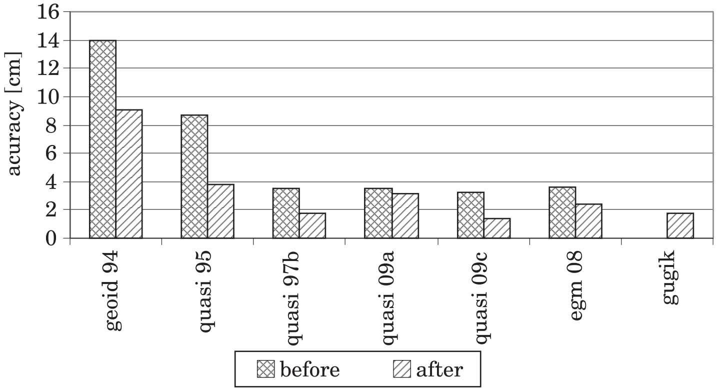


Fig. 1. Accuracy comparison of various geoid/quasigeoid models before and after fitting to the vertical reference system Kronsztad86 in Poland

Figure 1 represents the accuracy evaluation of various geoid/quasigeoid models before and after fitting to the vertical reference system Kronsztadt86. From figure 1 it can be seen that in the case of *geoid94* model if we apply the parametric model #3 then *geoid94* can be fitted to the vertical reference system Kronsztadt86 with an accuracy of ±9.1 cm.

If to the *quasi95* (= *quasi96*) model we apply the parametric model #3, then the fit to the vertical reference system is with accuracy ±3.8 cm. On the other hand *quasi97b* model was fitted to the vertical reference system by the collocation method and surprisingly good accuracy of the fitting ±1.0 cm was obtained (ŁYSZKOWICZ 2000).

Evaluation of accuracy of the fitted *quasi09a* model (parametric model#3) at the points of the POLREF network gives a mean error ±3.1 cm (ŁYSZKOWICZ 2009c). Accuracy of the *quasi09c* model computed from gravimetric data and geopotential model EGM08 by the collocation method after fitting to the vertical reference height system by the least squares collocation method, performed in this paper gives a mean error on the level of ±1.4 cm.

Accuracy of the fitting of EGM08 model to the vertical reference system Kronsztadt86 with the utilization of parametric model #3 is the same as by the collocation method and is equal to ±2.4 cm.

All accuracy evaluations given here were determined on the basis of the POLREF network. Investigations of the control traverse show, somewhat better accuracy. For example EGM08 model fits to the POLREF network with an accuracy ±2.4 cm while the fit to the points of the control traverse is ±1.9 cm.

Model “*GUGiK 2001*” approved in 2001 by the Head Office of Geodesy and Cartography is already a model which fits to the vertical reference system and therefore its new fitting was not done in the present paper leaving in the suitable place of figure 1 the value zero.

Geoid, gravity anomalies and deflections of the vertical from geopotential models

Information about the gravity field can be obtained from direct gravimetric measurements or computed from geopotential models. The first geopotential model up to degree and order 8 was elaborated by Żongołowicz in 1956 and characterized by low accuracy e.g. geoid heights could be computed with an accuracy of ±8 meters. The most recent geopotential model EGM08 enables to compute the geoid with the accuracy of few centimetres.

The aim of the preset section is to show how the accuracy of models: GEM-10B to the degree and order 36 (LERCH et al. 1978), OSU81 to the degree and order 180 (RAPP 1981), OSU91 to the degree and order 360 (RAPP et al. 1991), EGM08 to the degree and order 2100 (PAVLIS et al. 2008) during the last 50 years increased significantly.

The quasigeoid heights can be computed from the general formula

$$\zeta(r,\ \phi,\ \lambda) = \zeta_0 + \frac{\mathrm{GM}}{r\gamma}\sum_{n=2}^{n_{\max}}\left(\frac{a}{r}\right)^n\sum_{m=0}^{n}(C_{nm}\cos m\lambda + S_{nm}\sin m\lambda)\,P_{nm}\,(\sin\phi) \quad (12)$$

where C_{nm}, S_{nm} are fully normalized spherical harmonic coefficients of degree n and order m, $n_{\max}$ is the maximum degree of geopotential model, GM is the product of the Newtonian gravitational constant and mass of the geopotential model, r, ϕ, λ are spherical polar coordinates, a is the equatorial radius of the geopotential model and P_{nm} are the fully normalized associated Legendre functions. The term ζ_0 is the zero degree term due to the difference in the mass of the Earth used in IERS Convention and GRS80 ellipsoid. Detailed description of the term ζ_0 is given in the paper e.g. (ŁYSZKOWICZ 2009a).

In the present work in an analogous way, gravity anomalies Δg^{GM} and deflection of the vertical ζ^{GM}, η^{GM} were computed from various geopotential models. All computations were done using the program *geoco*l (TSCHERNING et al. 1992).

Quasigeoid heights were computed at 360 points of the POLREF network for the various geopotential models and were compared with "true" heights according to the formula (3) and then empirical standard deviations were computed for each model (fig. 2). From figure 2 it results that the accuracy of ζ computed from geopotential models increased extremely (almost 40 times) and presently without difficulties ζ in the area of Poland can be computed with accuracy of ±3.6 cm.

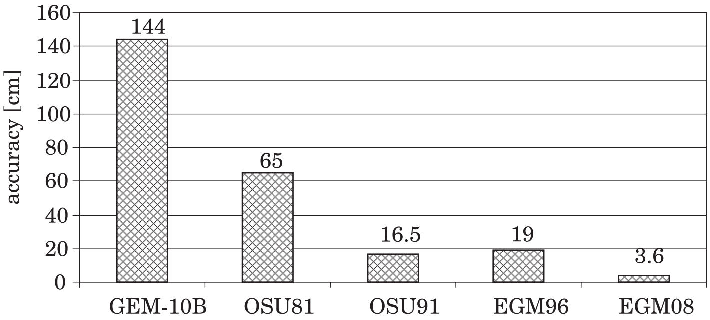


Fig. 2 Accuracy estimation of ζ from the various geopotential models

Terrestrial gravity data from the territory of Poland, used to evaluate the accuracy of gravity anomalies computed from geopotential models, consists of 147 530 mean gravity anomalies computed in a 1' × 1' grid. Gravity anomalies computed from the geopotential model were then compared with anomalies from terrestrial measurement, equation (4), and on the basis of such computed differences their accuracy was estimated (fig. 3). From that figure it can be concluded that at present, gravity anomalies in the area of Poland, can be computed from the EGM08 model with an accuracy of ±25 μm s^{-2}.

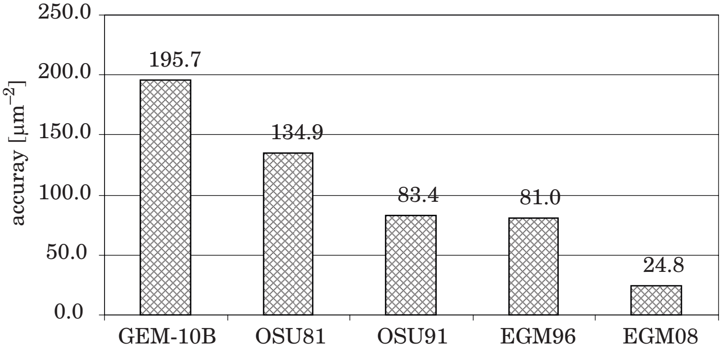


Fig. 3 Accuracy estimation of Δg from the successive geopotential models

171 astrogeodetic deflection of the vertical were used to estimate the accuracy of the deflections of the vertical ξ and η computed from the geopotential models. According to formula (5) appropriate differences were created and on their basis the accuracy was evaluated. From fig. 4 it can be seen that the accuracy of deflections of the vertical computed from geopotential models increase significantly and in of case the last EGM08 model they are ±0.6'' and ±0.7'' for the component ξ and η respectively.

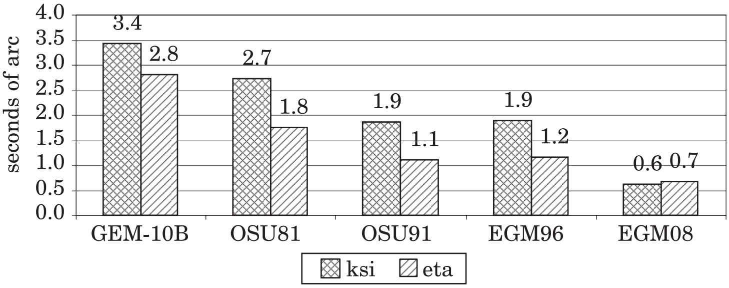


Fig. 4. Accuracy estimation of ξ η from the various geopotential models

Gravimetric deflections of the vertical

Traditionally deflections of the vertical are determined at the points of triangulation network from astronomical observations. The access to gravity data, which appeared in the last decade of the XX century, makes possible the computation of deflection of the vertical from Vening-Meinesz formulas.

First computation of gravimetric deflections of the vertical for the territory of Poland was done for the requests of the Head Office of Geodesy and Cartography in 1966 (ŁYSZKOWICZ 1996a). Deflections of the vertical ξ and η were computed from available at that time gravity anomalies[5] and geopotential model OSU91 by the fast Fourier method in a 1.5' × 3.0' grid. In order to assess their accuracy from the gridded data 171 gravimetric deflections of the vertical were interpolated at the points of astrogeodetic network and then the differences were computed according to formula (5) and their accuracy was estimated. This comparison gives mean errors ±0.59'' and ±0.47'' for components ξ and η respectively.

A second computation of gravimetric deflections of the vertical was realized in 2003 (ŁYSZKOWICZ 2003). In this computation the same gravity data set was used as in previous computation but the new and better geopotential model EGM96 available at that time. These calculations were made using Vening-Meinesz formulas, which were evaluated by the fast Fourier transform. The accuracy of compute deflections of the vertical is ±0.55'' and ±0.47'' for the component ξ and η respectively.

[5] At least 130 000 mean and point Faye anomalies.

Summary and conclusions

In the last years many geoid/quasigeoid models were computed for the territory of Poland. These models were computed using different methods and from various types of data. The computed geoid models i.e. astrogeodetic models and gravimetric models were computed using fast Fourier transform or the collocation method while geopotential models were computed from spherical harmonics coefficients.

Accuracy evaluation of these models was done on the basis of satellite networks such as POLREF network and control traverse, the points of which were connected to the national levelling network assuming that heights $N^{\mathrm{gps/levelling}}$ are error-free.

As a result it was found that absolute accuracy of the astrogeodetic geoid model is ±7 cm, accuracy of gravimetric geoid models *quasi97b* and *quasi09a* is ±3.5 cm, while the accuracy of the quasigeoid model computed by collocation is ±3.2 cm. Quasigeoid model computed from the new geopotential EGM08 model has an accuracy ±3.6 cm. The accuracy of the mentioned quasigeoid models tested on points of the control traverse is much better and is at the level of ±2.5 cm.

By suitable fitting of the quasigeoid model to the national vertical system one can get the surface, which is obviously not an equipotential surface and in the literature is called the correction surface. Accuracy of such correction surface in the case of *quasi09c* model is ±1.4 cm, while in the case of quasigeoid from geopotential model EGM08 is ±2.4 cm. Accuracy of the model "*GUGiK 2001*", which is of course fitted to the vertical system Kronsztadt86 is ±1.8 cm. It means that the above mentioned models have almost the same accuracy, but the model *quasi09 c* is the most accurate model.

If we reject the hypothesis that the term $N^{\mathrm{gps/levelling}}$ is error-free, then the accuracy of these geometrical distances depends on the accuracy of height h from satellite measurements and on the accuracy of connection to the vertical reference system. Accepting, that the accuracy of computation of h in the POLREF network is ±1 ... ±1.5 cm (ZIELIŃSKI et al. 1997) and, that the POLREF network was connected to the vertical reference system with the accuracy of ±1.5 cm (WYRZYKOWSKI 1998), (GELO 1994) it means that the quantity $N^{\mathrm{gps/levelling}}$ is determined with an accuracy of ±2.1 cm, what further investigations show is a very optimistic opinion. From these evaluations it appear that the POLREF network is not very suitable to estimate the accuracy of recent quasigeoid models and new testing networks are indispensable.

This new network could be the ASG-EUPOS network with height component determined on the level of a few millimetres and connected to the vertical reference system. It should ensure credible evaluation of various geoid/quasigeoid models and their proper fit to the vertical system (various possibilities).

Additionally in the present work it was proved that the accuracy of gravity anomalies computed from EGM08 model is of the order of ±25 μm s^{-2} and the accuracy of the deflections of the vertical is 0.6'' – 0.7'', what can replace difficult astronomic field measurements in many cases.

Translated by AUTHORS

Accepted for print 19.05.2012

References

BOKUN J. 1961. *Astrogravimetric Geoid Determination Referred to Krassovski's ellipsoid in the Area of Poland* (in Polish), Proceedings of the Institute of Geodesy and Cartography, VIII(1).

FORSBERG R. 1998. *Geoid Tayloring to GPS – with Example of 1-cm Geoid of Denmark.* Reports of the Finnish Geodetic Institute, 98:4.

GELO S. 1994. *Pismo Ministerstwa Gospodarki Przestrzennej i Budownictwa.* Warszawa.

Katalog względnych odchyleń pionu w Polsce. 1981. Instytut Geodezji i Kartografii, Warszawa.

KOTSAKIS C. 2008. *Transforming ellipsoidal heights and geoid undulations between different geodetic reference frames*, Journal of Geodesy, 82: 249–260.

KRYŃSKI J. 2007. *Precyzyjne modelowanie quasi-geoidy na obszarze Polski – wyniki i ocena dokładności.* Instytut Geodezji i Kartografii, seria monograficzna nr 13.

KRYŃSKI J., CISAK J., FIGURSKI M., MAŃK M., BIENIEWSKA H., MOSKWIŃSKI M., SĘKOWSKI M., ZANIMONSKIY Y., ŻAK Ł. 2005. *Przeprowadzenie pomiarów GPS oraz ewentualnych niezbędnych nawiązań niwelacyjnych na trawersach kontrolnych i opracowanie wyników.* Instytut Geodezji i Kartografii, Raport dla Instytutu Geodezji i Kartografii, Warszawa.

LEMOINE F.G., KENYON S.C., FACTOR J.K., TRIMMER R.G., PAVLIS N.K., CHINN D.S., COX C.M., KLOSKO S.M., LUTHCKE S.B., TORRENCE M.H., WANG Y.M., WILLIAMSON R.G., PAVLIS E.C., RAPP R.H., OLSON T.R. 1998. *The Development of the Joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) Geopotential Model EGM96.* NASA Technical Paper NASA/TP1998206 861, Goddard Space Flight Center, Greenbelt, Maryland.

LERCH F.J., WAGNER C.A., KLOSKO S.M., BELOTT R.P., LAUBSCHER R.E., RAYLOR W.A. 1978. *Gravity Model Improvement Using Geos3 Altimetry (GEM10A and 10B).* 1978 Spring Annual Meeting of the American Geophysical Union.

ŁYSZKOWICZ A. 1993. *The Geoid for the Area of Poland.* Artificial Satellites, 28(2), Planetary Geodesy, 19: 75–150.

ŁYSZKOWICZ A. 1994. *Opis algorytmu badania przebiegu geoidy na obszarze Polski, dane grawimetryczne i wysokościowe, grawimetryczna baza danych.* Raport No 11, Centrum Badań Kosmicznych PAN.

ŁYSZKOWICZ A. 1995. Raport z realizacji projektu badawczego KBN nr S605 014 04 „Problemy oceny dokładności przebiegu geoidy wyznaczonej metodą kolokacyjno-całkową przy zastosowaniu metody Fast Fourier Transform (FFT)". Warszawa.

ŁYSZKOWICZ A. 1996a. Sprawozdanie techniczne z realizacji umowy nr 17/CBK/96 między Departamentem Katastru, Geodezji i Kartografii a CBK PAN dotyczącej wyznaczenia składowych odchyleń pionu i odstępów quasi-geoidy od geocentrycznej elipsoidy GRS80 w 6852 punktach podstawowej osnowy geodezyjnej. Centrum Badań Kosmicznych PAN, Warszawa.

ŁYSZKOWICZ A. 1997. Raport z realizacji umowy 38/CBK/97 między Głównym Urzędem Geodezji i Kartografii a CBK PAN dotyczący utworzenia dla potrzeb GUGiK systemu obliczania odstępów quasi-geoidy model QUASI96 od elipsoidy GRS80 dla obszaru Polski. Warszawa.

ŁYSZKOWICZ A. 1998. Raport z realizacji projektu zamawianego KBN 008-07 „Założenia naukowe i metodyczne modernizacji krajowego układu wysokościowego". Warszawa, maj 1998.

Łyszkowicz A. 2000. *Improvement of the Quasigeoid model in Poland by GPS and levelling data.* Artificial Satellites, Journal of Planetary Geodesy, 35(1): 3–8.

Łyszkowicz A. 2003. *Gravimetric vertical deflections for the area of Poland.* Artificial Satellites, Journal of Planetary Geodesy, 38(4): 107–118.

Łyszkowicz A. 2009a. *Assessment of accuracy of EGM08 model over the area of Poland.* Technical Reports, 12: 118–134.

Łyszkowicz A. 2009b. *Badanie wpływu uwzględnienia odchyleń pionu na jakość grawimetrycznej quasi-geoidy na obszarze Polski.* Raport z realizacji projektu badawczego N N526 2163 33, Olsztyn.

Łyszkowicz A. 2010a. *Quasigeoid for the area of Poland computed by least squares collocation.* Technical Sciences, 13.

Łyszkowicz A. 2010b. *Refined astrogravimetric geoid in Poland.* Part I. Geomatics and Environmental Engineering, 4/1.

Łyszkowicz A. 2010c. *Refined astrogravimetric geoid in Poland.* Part II. Geomatics and Environmental Engineering, 4/2.

Łyszkowicz A., Denker H. 1994. *Computation of Gravimetric Geoid for Poland Using FFT.* Artificial Satellites, Planetary Geodesy, 21: 1–11.

Łyszkowicz A., Forsberg R. 1995. *Gravimetric Geoid for Poland Area Using Spherical FFT.* IAG Bulletin d'Information N.77, IGES Buletin N.4, Special Issue, Milano, pp. 153–161.

Pavlis N.K., Holmes S.A., Kenyon S.C., Factor J.K. 2008. *An Earth Gravitational Model to Degree 2160: EGM2008.* Presented at the 2008 General Assembly of the European Geosciences Union, Vienna, Austria, April 13–18.

Pażus R. 2001. *Instrukcja Techniczna G-2, Szczegółowa pozioma i wysokościowa osnowa geodezyjna i przeliczanie współrzędnych między układami.* Główny Geodeta Kraju, wydanie 5 zmienione, GUGiK, Warszawa.

Pażus R., Osada E., Olejnik S. 2002. *Geoida niwelacyjna 2001.* Geodeta, 5.

Rapp R.H. 1981. *The Earth's gravity field to degree and order 180 using SEASAT altimeter data, terrestrial data and other data.* OSU Report No 322.

Rapp R.H., Wang Y.M., Pavlis N.K. 1991. *The Ohio Stale 1991 geopotential and sea surface topography harmonic coefficient models.* Report No. 410, Department of Geodetic Science and Surveying. The Ohio Stale University, Columbus, Ohio.

Rogowski J.B., Bogusz J., Kujawa L., Kłęk M. 2005. *Opracowanie metody i utworzenie modelu geoidy astronomiczno-geodezyjnej.* Politechnika Warszawska, Raport dla Instytutu Geodezji i Kartografii, Warszawa.

Rogowski J.B., Kłęk M. 2003. *Ujednolicenie istniejących danych astronomiczno-geodezyjnych z wprowadzeniem do bazy danych.* Politechnika Warszawska, Raport dla Instytutu Geodezji i Kartografii, Warszawa.

Tanni L. 1949. *The regional rise of the geoid in Central Europe.* Annales Academiae Scientiarum Fennicae, Seria A, Helsinki.

Tscherning C., Forsberg R., Knudsen P. 1992. *The GRAVSOFT package for geoid determination.* First Continental Workshop On The Geoid In Europe "Towards a Precise Pan-European Reference Geoid for the Nineties" Prague, May 11–14.

Wyrzykowski T. 1988. *Monografia krajowych sieci niwelacji precyzyjnej I klasy.* Instytut Geodezji i Kartografii, Warszawa.

Zieliński J.B., Łyszkowicz A., Jaworski L., Świątek A., Zdunek R., Gelo S. 1997. *Polref-96 the New Geodetic Reference Frame for Poland.* Springer, International Association of Geodesy Symposia, Symposium 118: Advances in Positioning and Reference Frames, IAG Scientific Assembly, Rio de Janeiro, Brazil, September 3–9, pp. 161–166.

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

SEAMLESS COMMUNICATION FOR CRISIS MANAGEMENT

***Wojciech Wojciechowicz*[1,13], *Jacques Fournier*[2], *Miroslav Konecny*[3], *Stefan Vanya*[3], *John Stoodley*[4], *Phil Entwisle*[4], *Daniel M. Hein*[5], *Aurel Machalek*[6], *Apostolos Fournaris*[7], *Mikel Uriarte*[8], *Oscar Lopez*[8], *Shaun O'Neill*[9], *Hans Bradl*[10], *Zoltan Balogh*[11], *Emil Gatial*[11], *Ladislav Hluchy*[11], *Tomasz Mirosław*[12], *Jan Zych*[1]**

[1] ITTI sp. z o.o., [2] CEA-LETI Minatec, Gardanne, France, [3] Ardaco a.s., [4] QinetiQ Ltd., [5] Institute of Applied Information Processing and Communication, Graz University of Technology, [6] University of Luxembourg, [7] University of Patras, [8] NEXTEL S.A., [9] British Association of Public Salety Communications Officials, [10] Infineon Technologies AG, [11] Institute of Informatics, Slovak Academy of Sciences, [12] BUMAR sp. z o.o., [13] Institute of Computing Science, Poznań University of Technology

Key words: SECRICOM, Seamless communication, crisis management, Multi Bearer Router (MBR), Push To Talk (PTT), SECRICOM Silentel, Secure Docking Module (SDM), Secure Agent Infrastructure (SAI), Communication Security Monitoring and Control Centre (CSMCC), Seventh Framework Programme (FP7), Trusted Computing.

Abstract

SECRICOM – Seamless Communication for crisis management was a research and development project, realised within the Seventh Framework Programme (7PR). The aim of this project was to develop reference solution based on existing infrastructure, which will be capable to ensure secure and efficient communication for operational crisis management. The project was an answer to the European Security Research Advisory Board (ESRAB) report, in which key requirements for a communication system have been stated.

Secure and efficient communication system is a necessity for effective crisis management. It is assumed that such infrastructure may significantly increase rescue actions effectiveness. Currently, however, there are cases when various services (not only domestically but also internationally) use heterogeneous telecommunications systems. It results in the lack of or significant problems with mutual communication. Such situation is often considered problematic and posing a threat to the effective rescue actions.

For this purpose, a secure and multi-platform communications system (SECRICOM Silentel) has been developed within SECRICOM project. The Multi Bearer Router (MBR) optimise the backbone network by the use of multiple bearers and dynamic adjustment to various conditions. Advance mechanisms enhancing end-user devices' security – Secure Docking Module (SDM) – have been developed using Trusting Computing principles. Secure Agent Infrastructure (SAI) ensures – based on agents' infrastructure – secure access to distributed data. The system is supplemented with network monitoring platform – Communication Security Monitoring and Control Centre.

The SECRICOM project resulted in a communication system prototype, which is capable of ensuring interoperability as well as secure and efficient communication for operational crisis management. This system has been demonstrated on several occasions to the stakeholders.

PONADSYSTEMOWA ŁĄCZNOŚĆ DO ZARZĄDZANIA KRYZYSOWEGO

***Wojciech Wojciechowicz*[1,13], *Jacques Fournier*[2], *Miroslav Konecny*[3], *Stefan Vanya*[3], *John Stoodley*[4], *Phil Entwisle*[4], *Daniel Hein*[5], *Aurel Machalek*[6], *Apostolos Fournaris*[7], *Mikel Uriarte*[8], *Oscar Lopez*[8], *Shaun O'Neill*[9], *Hans Bradl*[10], *Zoltan Balogh*[11], *Emil Gatial*[11], *Ladislav Hluchy*[11], *Tomasz Mirosław*[12], *Jan Zych*[1]**

[1] ITTI sp. z o.o., [2] CEA-LETI Minatec, Gardanne, France, [3] Ardaco a.s., [4] QinetiQ Ltd., [5] Institute of Applied Information Processing and Communication, Graz University of Technology, [6] University of Luxembourg, [7] University of Patras, [8] NEXTEL S.A., [9] British Association of Public Salety Communications Officials, [10] Infineon Technologies AG, [11] Institute of Informatics, Slovak Academy of Sciences, [12] BUMAR sp. z o.o., [13] Instytut Informatyki, Politechnika Poznańska

Słowa kluczowe: SECRICOM, „bezszwowa" komunikacja, zarządzanie kryzysowe, Multi Bearer Router (MBR), Push To Talk (PTT), SECRICOM Silentel, Secure Docking Module (SDM), Secure Agent Infrastructure (SAI), Communication Security Monitoring and Control Centre (CSMCC), siódmy program ramowy (7PR), Trusted Computing.

Abstrakt

SECRICOM – Seamless Communication for crisis management to projekt badawczo-rozwojowy, który został zrealizowany w ramach siódmego programu ramowego (7PR). Celem projektu było wypracowanie bezpiecznej i, co ważne, bazującej na istniejącej infrastrukturze platformy komunikacyjnej do operacyjnego zarządzania kryzysowego. Projekt ten stanowi odpowiedź na raport European Security Research Advisory Board (ESRAB), w którym określono najważniejsze wymagania odnośnie do systemu komunikacji.

Bezpieczny i wydajny system komunikacji jest warunkiem koniecznym do efektywnego zarządzania w sytuacjach kryzysowych. Przyjmuje się, że taka platforma jest w stanie znacząco zwiększyć efektywność prac służb ratunkowych. Obecnie jednak są przypadki, gdy służby ratunkowe (nie tylko na arenie międzynarodowej, lecz także podczas działań w jednym kraju) korzystają z niejednorodnych systemów telekomunikacyjnych, co często skutkuje brakiem lub istotnymi problemami z wzajemną łącznością. Sytuacja ta jest postrzegana jako problematyczna i stanowi zagrożenie dla efektywnego działania służb ratunkowych.

W ramach projektu SECRICOM opracowano system międzyplatformowej, bezpiecznej łączności SECRICOM Silentel. Za optymalizację transmisji danych (w tym wykorzystanie wielu nośnych oraz dynamiczne dostosowywanie się do warunków) w sieci dystrybucyjnej oraz szkieletowej odpowiada Multi Bearer Router (MBR). Zaawansowane mechanizmy zwiększające bezpieczeństwo urządzeń końcowych – Secure Docking Module (SDM) – opracowano z wykorzystaniem pryncypiów Trusted Computing. Secure Agent Infrastructure (SAI) zapewnia – oparty na infrastrukturze agentów – bezpieczny dostęp do rozproszonych danych. System uzupełnia platforma nadzoru nad siecią – Communication Security Monitoring and Control Centre.

Jako rezultat projektu zbudowano oraz kilkukrotnie zademonstrowano prototypową wersję systemu komunikacji. System ten jest zdolny do zapewnienia interoperacyjnej, bezpiecznej i wydajnej łączności w zarządzaniu w sytuacjach kryzysowych.

The SECRICOM Project

The SECRICOM project was a FP7 collaborative and integration research project, addressing the Security Theme in Call FP7-SEC-2007-1 in Topic SEC-2007-4.2-04 Wireless communication for EU crisis management. The main aim of the project was to create Seamless Communication for Crisis Management for EU Safety.

The project was started in September 2008 and finished – as planned – after 44 months (April 2012). The budget was 12.468.847 (incl. 8.606.791 co-funded from FP7 programme) and the project have been realised by 14. partners from:

– **Industry** – QinetiQ (project coordinator), BUMAR, Hitachi, Infineon.

– **SME** – Ardaco (technical coordinator), CEA-LETI, Geothermal Anywhere, iTTi, Nextel.

– **University** – Universite du Luxembourg, Institute of Informatics, Slovak Academy of Sciences, Graz University of Technology, University of Patras.

– **End-user** – British APCO.

Figure 1 presents the structure of the project.

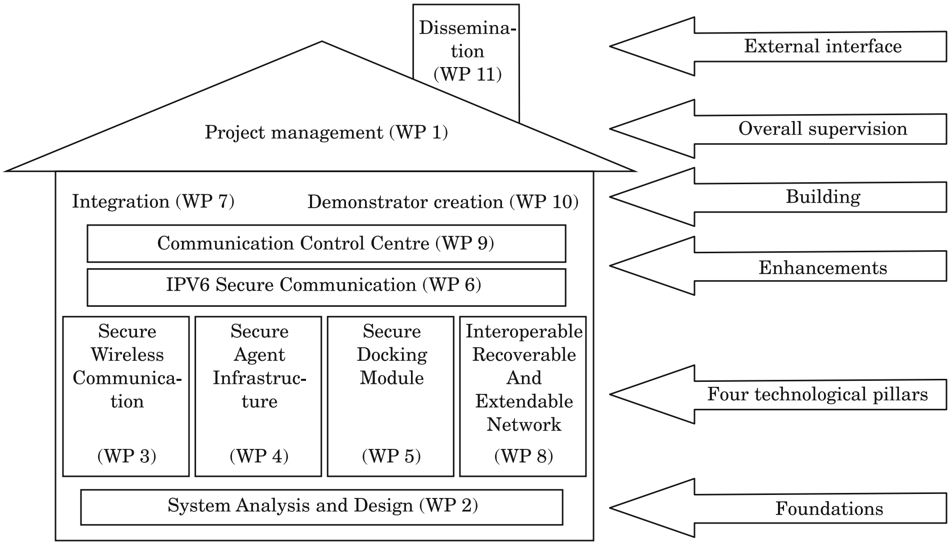

Dissemination (WP 11)
Project management (WP 1)
Integration (WP 7)
Demonstrator creation (WP 10)
Communication Control Centre (WP 9)
IPV6 Secure Communication (WP 6)
Secure Wireless Communication (WP 3)
Secure Agent Infrastructure (WP 4)
Secure Docking Module (WP 5)
Interoperable Recoverable And Extendable Network (WP 8)
System Analysis and Design (WP 2)
External interface
Overall supervision
Building
Enhancements
Four technological pillars
Foundations

Fig. 1. Project structure
Source: SECRICOM Grant Agreement Annex I – Description of Work.

Motivation

The SECRICOM project has been established in response to the European Security Research Advisory Board (ESRAB) report (published in September 2006) in which the key requirements of new communication infrastructure have been proposed:

– Secure in terms of protection against tapping and external intrusion.

– Trusted in the sense of behaving as expected by meaning minimising the threats for failure to be a basis for creating emergency solutions.

– Providing enhanced connectivity between various networks and devices.

– Ensuring transmission of different data types such as multimedia (e.g.: voice, picture, video sequences), geopositioning, etc.

– Supporting advanced search functions embedded in the infrastructure itself (Meeting the challenge.)

The above mentioned requirements are in line with the current market needs, including:

– Growing demand for seamless, resilient and secure communication among various emergency communication systems required by public safety organizations and their users.

– Necessity of ubiquitous multimedia communication, available everywhere, coupled with real-time access to relevant information in the crisis area.

– Need for recoverability and alternative restoration of damaged communication cells for infrastructure functionality.

– Requirement of authorization, authentication, data protection against misuse, quick and flexible data acquisition whilst retaining Push To Talk simplicity of operation.

– Interoperability and interconnectivity of existing communication platforms (SECRICOM Grant Agreement Annex I – Description of Work).

The SECRICOM project was intended to fulfil those requirements, and provide interoperable and efficient communication system dedicated to first responders. The main innovations areas of the SECRICOM projects are:

– Interconnectivity of commercial (e.g. GSM, UMTS, Citizen Band) and specialized communication systems (e.g. TETRA).

– Seamless and secure interoperability of existing mobile devices already deployed.

– Efficient multi-bearer network utilisation,

– Software layer based on mobile agents' paradigm.

– Security based on chip-level trusted module.

Concept

The SECRICOM project was aimed at a reference security platform development for EU crisis management operations with two essential ambitions:

1. Solve or mitigate problems of contemporary crisis communication infrastructures (e.g. TETRA, GSM, Citizen Band, IP) such as poor interoperability of specialized communication means, vulnerability against tapping and

misuse, lack of possibilities to recover from failures, inability to use alternative data carriers as well as high deployment and operational costs.

2. Add new smart functions to existing services which will make the communication more effective and useful for users. Smart functions will be provided by distributed IT systems based on an agents: infrastructure. Achieving these two project ambitions will allow creating a pervasive and trusted communication infrastructure fulfilling crisis management users requirements, ready for immediate application.

The SECRICOM solution was based on four technological pillars:

1. Secure and encrypted mobile communication based on existing infrastructures (e.g. TETRA, GSM, UMTS networks) – secure Push To Talk.

2. Improved interoperability among various existing communicating systems, creating recoverable networks with seamless connectivity.

3. Introduction of distributed systems and the agent paradigm forming a smart negotiating system for parameterization and independent handling of requests essential for immediate reaction.

4. Security based on trusted hardware enhancing the data confidentiality and the users privacy (SECRICOM Grant Agreement Annex I – Description of Work).

The SECRICOM infrastructure was designed mainly for crisis management communication (rescuers, fire brigades, special forces, police, healthcare, etc.), but during the project's live cycle also other potential end-users have been identified. The SECRICOM delivers an interface between selected systems currently deployed for crisis management and new generation systems which could be developed in next decade, such as SDR. An important goal is to enable seamless and secure interoperability between currently used radio systems. Achieving the latter will ensure that already invested resources are preserved; also, developments and emerging technologies can be used in the future.

The SECRICOM implementation principles were as follows:

- Provide value to end-users.
    - Learn from end users.
    - Provide new services and applications.
- Supplement existing technologies (not replace).
- Integrate with existing systems (like TETRA).
- Open interfaces to support extensibility (no vendor lock-in).
- Security built deep inside (not an afterthought).
- High availability/reliability.
    - Throughout testing.
    - Support of multiple bearers – MBR.
    - Graceful degradation and QoS.
    - On-site deployable infrastructure.

SECRICOM architecture

The communication system architecture is smart and innovative concept that allows technical interoperability, and in terms of this, is able to extend communications across different agencies and across different countries. The SECRICOM is also technically expandable, thus able to extend communications to places where it is usually not capable of achieving ubiquitous operations.

It is foreseen, that the features of SECRICOM project will impact directly to communications systems and communication networks for Emergency services, but the project's results and technology could be used by civilian markets as well.

SECRICOM system is based on following technologies:

– SECRICOM Silentel – a client-server communication system using IP protocol. It optimizes and protects the way teams of people communicate without being concerned about misuse of information. Regardless of whether device is used to communicate, the connection is secure and safe. The basis of the system is PTT technology, a two-way communication system, which works like a two-way radio, however, with the possibility of transfering voice and other data types (e.g. multimedia, text messages, control data, etc.).

– Secure Docking Module – in order to provide security for agents that dock on to a trusted agent network, the SECRICOM project proposes the usage of Secure layer based on hardware module so-called Secure Docking Module (SDM). The design of SDM is based on Trust Computing principles.

– Secure Agent Infrastructure – designed for mobile services with agent-like features (mobility, pro-activity) which would execute on secure devices. In general, it consists of interconnected trusted (TS) and untrusted servers (US). Each agent has features and "abilities", which are used for the enactment of certain processes. The processes enactment is designed for the management of crisis situations in which information collection from multiple untrustworthy environments is required.

– Communication Security Control Centre – collects information to assure the secure status of the SECRICOM system, presenting it at a fixed location.

– Multi Bearer Router – an intelligent adaptive routing device enabling seamless inter-networking in a multi-bearer, multi-node, mobile environment designed to optimise network performance whenever users operate in environments where connectivity is poor

– IPv6 – All the modules developed for SECRICOM are eligible to cope with an IPv6 environment. The modules like SECRICOM PTT, Secure Docking Module, Multi-Bearer Router and Communication Security Control Centre are capable of handling the IPv6. This protocol and its impact on secure communication was studied and described in details by University Luxembourg. The SECRICOM's compatibility with the IPv6 is confirmed by IPv6 Forum, where the IPv6 Silver Ready Logo have been awarded.

General SECRICOM network architecture is presented in the Figure 2.

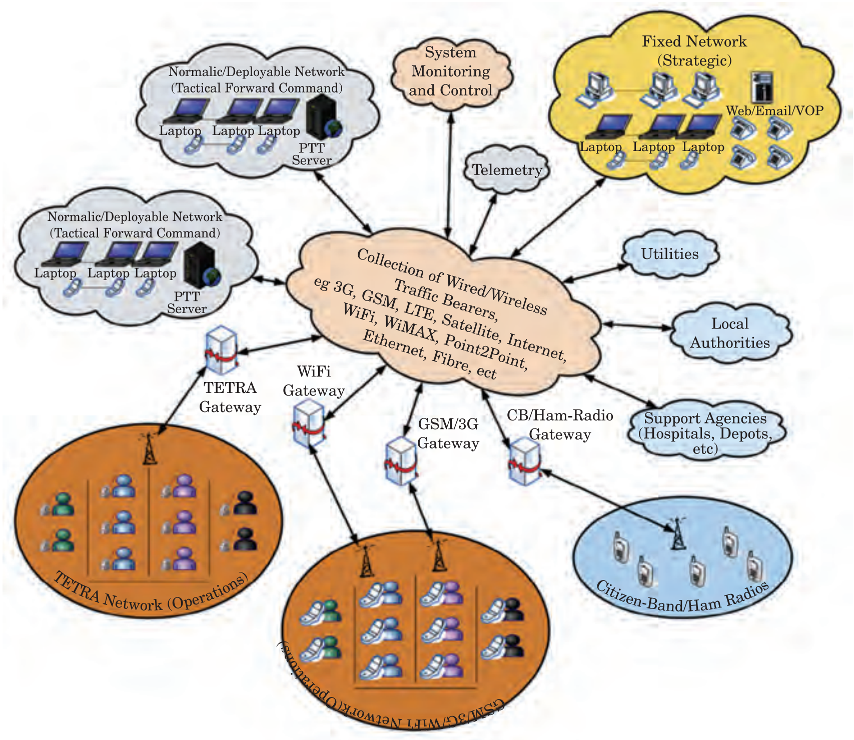

Normalic/Deployable Network
(Tactical Forward Command)
Laptop
Laptop
Laptop
PTT Server
System Monitoring and Control
Fixed Network (Strategic)
Web/Email/VOP
Laptop
Laptop
Laptop
Telemetry
Normalic/Deployable Network
(Tactical Forward Command)
Laptop
Laptop
Laptop
PTT Server
Collection of Wired/Wireless Traffic Bearers, eg 3G, GSM, LTE, Satellite, Internet, WiFi, WiMAX, Point2Point, Ethernet, Fibre, ect
Utilities
Local Authorities
TETRA Gateway
WiFi Gateway
GSM/3G Gateway
CB/Ham-Radio Gateway
Support Agencies (Hospitals, Depots, etc)
TETRA Network (Operations)
Citizen-Band/Ham Radios
GSM/3G/WiFi Network(Operations)

Fig. 2. SECRICOM Network Architecture
Source: SECRICOM Deliverable D2.1 – Analysis of external and internal system requirements.

Push To Talk (SECRICOM Silentel)

SECRICOM Silentel is a client-server communication system. The application has been designed to support the First Responders in their day-to-day missions, as well as critical situations. The main aim of this solution is to optimise and protect the communication between end users in a seamless way, using existing infrastructure (incl. end users terminals and network infrastructure). The communication between endpoints is encrypted to prevent any transmitted data (voice, text, images, position, status, etc) from misuse. The user requirements were defined in SECRICOM User Requirements[1]. The key features include:

1. Voice communication.
   a. One-to-one full duplex.
   b. One-to-many half-duplex group call (Push To Talk).
2. Online group management.
3. Instant text messaging,
4. Smart text messaging.
5. Data delivery.
6. Video communication.
7. User location information and mapping, based on GPS.

The system is based on the IPv4/IPv6, which facilitates the system's scalability as well as integration with other technologies. The solution's High Level Architecture is given in the Figure 3.

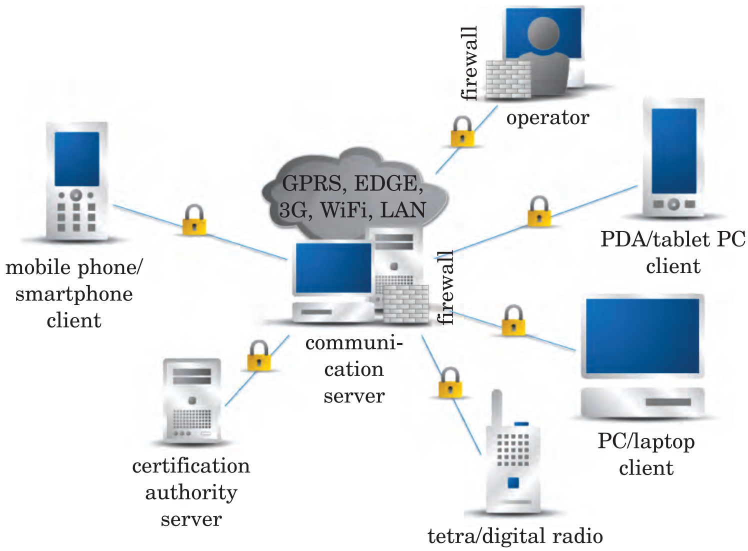


Fig. 3. Secricom Silentel (PTT) high level architecture
Source: ARDACO dissemination materials.

Main parts of SECRICOM Silentel architecture are as follows:

– Communication Server – switching centre module to provide communication services to all system's users.

– Certification Authority – trust module for server and users certification creation, validation and revocation.

[1] D2.1 – Analysis of external and internal system requirements.

– Operator Studio – user management tool (incl. personal contact list definition).

– End user application – software application to be deployed on end user's terminal. Currently several implementations on various operating systems are available – incl. Symbian, Windows Mobile, iOS, Windows and Android.

– SECRICOM gateways – device to provide interoperability with some legacy networks (e.g. TETRA, Citizen Band).

Multi-Bearer Router

The routers that are currently used in managing major incidents and crisis situations tend to be inflexible and inefficient when it comes to meeting the following requirements:

– Interfacing to selected communication systems which are taken into the field.

– Interfacing to a number of communication systems which have survived a major incident or were brought into the incident zone by a supporting agency.

They do not provide a sufficient level of resilience to enable the business processes and operations to continue to operate in areas where existing communication systems and infrastructure are either destroyed by an incident or poor in its original form.

The Multi-Bearer Router (MBR) technology has been designed to fill that gap. It is an intelligent adaptive routing device enabling seamless internetworking in a multi-bearer, multi-node, mobile environment designed to optimise network performance whenever users operate in environments where the connectivity is poor. There is a need for SECRICOM solution to seamlessly support different types of user traffic (with different QoS and security requirements) over different communication bearers (with a range of capabilities), depending on the end user environment (e.g. disaster relief with ad hoc communications, mobile working with dynamically changing access to communications service provider networks). Leveraging legacy communication systems is possible due to integrating together modern satellite communications, mature security, provisioning private networks, open architectures, application persistence and ubiquitous mobile broadband. Conceptually, the SECRICOM's MBR performs two functions:

– Allows creating an efficient network-of-networks as the basis for business focused traffic delivery (roaming-like) across different networks. This network-of-networks is achieved by transparently integrating available communication and network systems (both wireless and wired forms); and making intelligent and flexible traffic routing decisions based on multiple factors, e.g. user application, availability and reliability of networks, effective bandwidth, cost and security.

– Provides a gateway for SECRICOM end systems and devices to the created network-of-networks through a single Service Access Point (SAP).

The MBR is independent of the type of traffic and is intelligent enough to inhibit inappropriate data streams such as video, for example, over unsuitable bearers whilst still maintaining the bearer usage for other systems. The technology allows highly confidential data to always be routed over the TETRA bearer. It simultaneously routes less sensitive high bandwidth data, such as still photographs, video or CCTV images, over other high capacity bearers. If one of these bearers is lost, the application routes seamlessly over an alternative bearer, with no need for user intervention and no interruption to the service. The Multi-Bearer Router features a unique intelligent policy engine. This constantly monitors the constituent network capabilities and modifies traffic delivery policies dynamically according to pre-defined business needs and the level of services available. This enables the user to automatically use the available networks with the least impact on performance and without the risk of security compromise. The flexible policies are tailored to meet objectives such as operational imperatives, user needs and specific application requirements.

Key benefits include:

- Increases efficiency through provisioning of mobile broadband.
- Fast and reliable access to information for better situational awareness.
- Improves communication coverage.
- Prioritises business critical information flow.
- Supports minimal configuration in field.
- Provides unified access to all available communications services.

Secure Docking Module

Communication security is a well-established in today's networked computers. Security expert Eugene Spafford once said using secure communication technologies, such as SSL, is similar to "using an armored truck to transport rolls of pennies between someone on a park bench and someone doing business from a cardboard box". This refers to the communication channel being secure due to technologies like SSL and SSH, when used with strong cryptographic algorithms such as RSA, elliptic curve cryptography and AES. So what about the communication end-points? "Park benches" and "cardboard boxes" have no place in a secure communication infrastructure, such as the one developed by the SECRICOM project. This is the gap that the Secure Docking Module wants to close.

General purpose computing platforms such as PCs, laptops and mobile phones are all vulnerable to over-the-network software attacks. The enormous amount of known vulnerabilities in operating systems and applications is testament to this vulnerability. A wide variety of malicious software is known to exploit such vulnerabilities to take over unsuspecting computing platforms. Malicious software on a computing platform used in crisis management is unacceptable. For SECRICOM to be able to leverage the power of mobile computing platforms for crisis management, as for example with the Secure Agent Infrastructure, the security of the computing platforms must be established.

The Secure Docking Module is a platform-software-configuration verification device. When malicious software infects a computing platform, it usually integrates itself into the attacked platform. This integration changes the platform software configuration. Now, if an external arbiter could detect this change of software configuration, it might be able to deem the platform insecure for use within a crisis management infrastructure. Thus, spreading of the malicious software to other computing platforms in the network would be hindered. The Secure Docking Module is just such an arbitration device.

The Secure Docking Module works in conjunction with the Trusted Docking Station. The Trusted Docking Station uses recent security additions to commodity-of-the-shelf computing platforms to achieve two goals. First, the Trusted Docking Station provides the ability to measure and attest a platform's software configuration. Second, it establishes a strongly isolated execution environment for crisis management applications. Thus, the Trusted Docking Station is capable of reliably reporting its software configuration to the Secure Docking Module. In addition, it provides an execution environment where each software component is protected by virtualization technology form other components of the system.

The Secure Docking Module provides the cryptographic resources required to access the SECRICOM crisis management infrastructure. If a SECRICOM software component wants to connect to the SECRICOM network it reports the software configuration of its execution environment to the Secure Docking Module. The Secure Docking Module then proceeds to check the integrity and freshness of the platform software configuration report using cryptographic algorithms. If the integrity and freshness of the report are verifiable, the Secure Docking Module will validate the platform software configuration. Only if the platform software configuration is valid, does the Secure Docking Module provide its cryptographic resources to the Trusted Docking Station, and thus to the SECRICOM software component. In this way only platforms with a verified software configuration and approved applications are able to connect to the SECRICOM infrastructure.

Secure Agent Infrastructure

One of the challenges of the communication infrastructures for crisis management is to add new smart functions to existing services which would make the communication more effective and helpful for users. The aim is to provide smart functions via distributed information systems which should provide a secure distributed paradigm to achieve confidentiality and access to resources. In the SECRICOM project requirements, design and implementation of such distributed information system – called Secure Agent Infrastructure (SAI) was enacted.

Requirements for such infrastructure were to provide a smart negotiating system for parameterization and independent handling of access requests to achieve rapid reaction. By fulfilling these goals a pervasive and trusted communication infrastructure satisfying the requirements of crisis management authorities and ready for immediate application was introduced. SAI represents one of the core parts of the SECRICOM communication infrastructure.

In crisis situations there are requirements to collect information from legacy systems of various organizations and from human operators in order to semi-automatically manage the crisis mitigation process or to enact decisions at various management levels. This collection of information must be enacted in a secure manner while ensuring trust between both parties – information consumers and information providers. Many actors participate in a crisis situation. Information gathering is enacted by secure agents either from legacy systems or from human end-users through mobile devices. Agent technology was selected due to the ability to fulfill such requirements through support of mobile and dynamically deployable executable code.

Additionally agents require safe secured place to store sensitive information (such as cryptographic credentials) and provide interfaces to retrieve these keys, ways to attest a platform and provide interface to safely communicate with legacy systems – all these functionalities are provided to the agent platform by a hardware module called Secure Docking Module (SDM) which was also developed in scope of the Secricom projects.

The SDM allows agents to dock on a secure communication infrastructure by ensuring the state of the device it is supporting. The SAI is a distributed system and operates on confidential data. Therefore, the system must protect its integrity against data loss/theft and data modification. In a distributed system, data protection concerns are not limited to data transmission. As the data is processed in different physical computing platforms it must be established that all data processing entities adhere to the same security policy for the data.

The data security policy adherence is enforced by ensuring the software configuration of a computing platform before it is connected to the SECRICOM infrastructure. To this end the SDM protects communication keys and credential information and only releases this information to the host platform if this platform is in an approved software configuration. The process of establishing the fact that a platform has an approved software configuration is called local attestation verification.

Conceptually, the SDM protects a small set of key pairs for asymmetric cryptography, but in general is capable of protecting arbitrary data up to a specific size. The SDM's key protection facilities are a standard function, which could already be implemented with today's smart cards or hardware security modules. The SDM extends this standard function by only releasing these keys to a host device if and only if this host device is in a trusted state. This host device is called Trusted Docking Stations (TDS).

Communication Security Monitoring and Control Centre

The main purpose of the Communication Security Monitoring and Control Centre (CSMCC) is to provide Security Model, suitable for secure and interoperable communications under crisis, which could be applied in the SECRICOM communication infrastructure. The Security Model defines the properties, capabilities, processes and controls that a secure infrastructure should contain to protect against various threats.

Key features of SECRICOM Communication Security Monitoring and Control Centre:

– Increased protection of assets: various protection mechanisms which control access and usage policies, scalable network architecture, auditing tools, and security assurance monitoring.

– Improved threat detection: new traffic patterns and event management policies and correlation for anomalous events

– Enhanced reaction for hostile environments: increased network resilience by enhancing IT structure, traffic blocking and isolating as well as alternative routing

– Fast recovery for crisis critical communications: quick and efficient recovering plans and mechanisms. One of the main strengths and unique features of the CSMCC platform in SECRICOM is the set of custom agents that have been deployed along the communication infrastructure. These enhanced agents provide new detection and action capabilities, such as adaptive routing features in case of network failure or congestion and VoIP traffic monitoring.

Additionally, Security Middleware Services and Framework was developed to measure, document and maintain the security of SECRICOM services, which are based on telecommunication services.

Designing the communication infrastructure security monitoring and control centre, started with a risk assessment of the SECRICOM system. It consisted of a deep analysis of the operating system, in order to define key assets, identify their security vulnerabilities and should risk occur, evaluate its impact. The outcome of risk assessment was a set of security requirements that the SECRICOM security model fulfils to provide an effective security management. This has been supported by a team of users who updated and validated these security requirements. The analysis of the security requirements results in a number of countermeasures and security mechanisms that are used to mitigate the level of risk and protect the SECRICOM systems against any kind of security threats. Finally, all these security principles and guidelines are aggregated into the SECRICOM security model in order to ensure continuous security of communication infrastructure in a continuous way. The security model is supported by a security middleware named Communication Security Monitoring and Control Centre (CSMCC) that provides not only a collection of security information and security status monitoring capabilities, but also active control mechanisms. They provide enhanced protection, improved detection, faster reaction and stronger risk mitigation, more effective incident impact mitigation and quicker restoration.

Conclusions

In this article we considered the SECRICOM project as an answer to the interoperability problems between first responders' communication systems. Those problems stem from different communication networks used by various rescue services, low security level and thus high vulnerability to different threats. The SECRICOM system was designed to provide unified, secure and seamless communication between various end-users and to improve the effectiveness of their work during crisis situation.

This paper provided a survey of all the technologies used within the SECRICOM system. The SECRICOM project has successfully fulfilled all requirements and designed a prototype capable of:

– Exploiting & optimizing existing communication systems.

– Enhancing interoperability among heterogeneous secure communication systems.

– Enhancing interconnectivity between different networks and User Access Devices.

– Interfacing towards emerging SDR systems.

– Mitigating key capability gaps faced by users of existing systems.

Thus, the “Seamless Communication for Crisis Management” proof of concept has been achieved. The SECRICOM has also been successfully demonstrated to the stakeholders during national as well as European events, including:

– BAPCO 2010 in Business Design Centre, London, UK.

– Civil Protection NATO Seminar in Lest Training Village, Slovakia.

– BAPCO 2011 in Business Design Centre, London, UK.

– ASTER 2011 in Żagań, Poland.

– SECRICOM demonstration 2012 in Portsmouth Technology Park, UK.

where the positive feedback have been received.

Acknowledgments

This work was supported by the SECRICOM project (EC FP7-SEC-2007 grant 218123).

Translated by Authors

Accepted for print 30.06.2012

References

Fournaris A., Hein D., Scheibe M., Fournier J., Verdier M. 2010. *Design of the Secure Docking Module*. Infineon Technologies AG.

Gatial E., Balogh Z., Hluchý L. 2010. *Platform for distributed execution of agents for trusted data collection*. Procedia Computer Science, 1: 2023-2032.

Goban-Klas T., Sienkiewicz P. 1999. *Społeczeństwo informacyjne: Szanse, zagrożenia, wyzwania.* Wydawnictwo Postępu Telekomunikacji, Kraków.

Hluchý L., Balogh Z., Gatial E. 2010. *Distributed agent-based architecture for management of crisis situations using trusted code execution*. SAMI 2010 8th International Symposium on Applied Machine Intelligence and Informatics Proceedings, pp. 25–30.

Meeting the challenge: the European Security Research Agenda. European Security Research Advisory Board, September 2006 – http://ec.europa.eu/enterprise/policies/security/files/esrab–report–en.pdf

Scheibe M., Hein D., Reymond G., Fournier J., Fournaris A.P., Hudek V., Slovak L. 2011. *SECRICOM WP5 Design of the Chip and Emulator*. Technische Universität Graz.

SECRICOM D2.1 – Analysis of external and internal system requirements.

SECRICOM Dissemination materials.

SECRICOM Grant Agreement Annex I – Description of Work.

Wojciechowicz W., Zych J. 2011. *Koncepcja infrastruktury telekomunikacyjnej o podwyższonej niezawodności*. In: *Bezpieczeństwo współczesnego świata – Informatyka, technika i gospodarka*. Red. Z. Dziemianko, WSHiU Poznań.

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

USER REQUIREMENTS FOR MISSION-CRITICAL APPLICATION – THE SECRICOM CASE

***Shaun O'Neill*[1], *Jim Strother*[1], *Jan Zych*[2], *Wojciech Wojciechowicz*[2,3]**

[1] British Association of Public Salety Communications Officials
[2] ITTI sp. z o.o., Poznań
[3] Institute of Computing Science, Poznań University of Technology

Key words: SECRICOM, Interoperability, Crisis Management, Requirements Engineering, Critical system

Abstract

The SECRICOM Project as a communication system for operational crisis management, requires paying significant attention to the requirements engineering phase. Any mistakes made during the requirements gathering phase may affect the subsequent software development phases, which creates excessive operational risks for the users of the system. These types of risks – as in any other critical systems – could have serious consequences, such as inefficiency of rescue actions and loss of lives.

This article presents the requirements engineering process, which was defined and carried out for the needs of the SECRICOM project. It describes the system's environment (the crisis management reference structure and the main organizational rules) and its impact on the developed. As a result, a requirements engineering process for SECRICOM is proposed. Finally, main points of gathered requirements are presented.

WYMAGANIA UŻYTKOWNIKA DOTYCZĄCE SYSTEMU KRYTYCZNEGO – PRZYPADEK SECRICOM

***Shaun O'Neill*[1], *Jim Strother*[1], *Jan Zych*[2], *Wojciech Wojciechowicz*[2,3]**

[1] British Association of Public Salety Communications Officials
[2] ITTI sp. z o.o., Poznań
[3] Insytut Informatyki, Politechnika Poznańska

Słowa kluczowe: SECRICOM, interoperacyjność, zarządzanie kryzysowe, inżynieria wymagań, system krytyczny.

Abstrakt

W systemie SECRICOM, ze względu na tworzenie systemu komunikacji do operacyjnego zarządzania kryzysowego, szczególnie ważne było położenie szczególnego nacisku na etap gromadzenia wymagań. Błędy popełnione na etapie specyfikacji wymagań mogą rzutować na kolejne etapy wytwarzania systemu, co w rezultacie generuje nadmiarowe ryzyka dla użytkowników systemu. Ryzyka te – jak w przypadku innych systemów krytycznych – mogą spowodować poważne konsekwencje, w tym obniżenie skuteczności akcji ratunkowych, a nawet straty po stronie ludności.

W artykule przedstawiono proces inżynierii wymagań, który zdefiniowano oraz przeprowadzono na potrzeby projektu SECRICOM. Przedstawiono środowisko systemu (zarówno referencyjną strukturę zarządzania kryzysowego, jak i główne zasady organizacji) oraz określono wpływ na budowany system. Na zakończenie przedstawiono główne wnioski z zebranych wymagań.

Introduction

This paper describes an approach to manage user requirements for the mission critical application, which was developed within the SECRICOM project. The main aim of the SECRICOM project is to propose a seamless and secure reference communication platform for EU crisis management operations. For that purpose, the first step within the SECRICOM project was to specify the requirements to be fulfilled by such a platform. As a result, not only user requirements were specified, but also a methodology for managing user needs for mission-critical application was created.

This article consists of four chapters. In the first chapter, the crisis management structure used by SECRICOM to gather the requirements is described. The second chapter explains the principles of crisis management and its impact on the capability gap analysis. The third chapter presents selected user requirements gathered in the SECRICOM project. The conclusions are provided in the fourth chapter.

Reference crisis management structure

In order to better address the user requirements, a reference crisis management structure has been proposed within the SECRICOM project and it is going to be described in this section. Typically, operational emergency services use a 3-tier command structure. For that reason the same command chain was proposed for the aim of collecting user requirements' in SECRICOM. Members of senior civil protection, and emergency service personnel in the UK, Luxembourg, Slovakia, Spain, Sweden and Poland, validated the structure. All the aforementioned parties agreed that the basic operational structure for the emergency services is similar, and it is based on three tiers:

strategic command, tactical command and operational (called also ground command). In the UK these tiers are represented by three types of metals – Gold, Silver and Bronze – hence the colours in Figure 1.

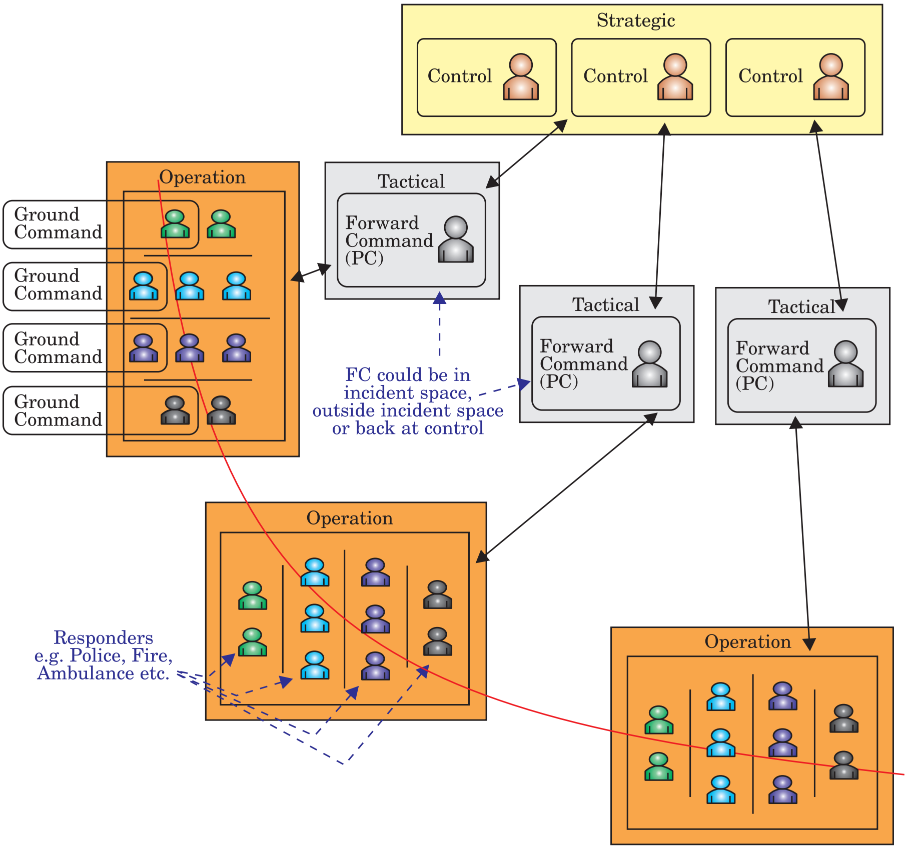


Fig. 1. Emergency Service Command Structure
Source: SECRICOM project.

The command structure has significant impact on gathering the user requirements; not only in terms of functional requirements (e.g. circuit-switched voice communication, Push To Talk, data transfer) but also non-functional requirements (e.g. security requirements) – as confirmed during the project.

Principle of crisis management

Crisis management is a process carried out by public authorities in co-operation with organisations, institutions and society in general, to cope with the crisis and ensure broadly defined public safety. Assuming that crisis management includes the whole sphere of the crisis situations, prevention issues occurring before, during and after an event, it is taken that its process should comprise of four permeating and highly connected fields of action that all constitute four phases of the crisis management. In each phase, challenges related to communication are identified. What's more, each of them expects different functionalities from communication. It seems that synthetic characterisation of each phase is necessary. Decision making process within the crisis management is greatly determined by the quality of system communication.

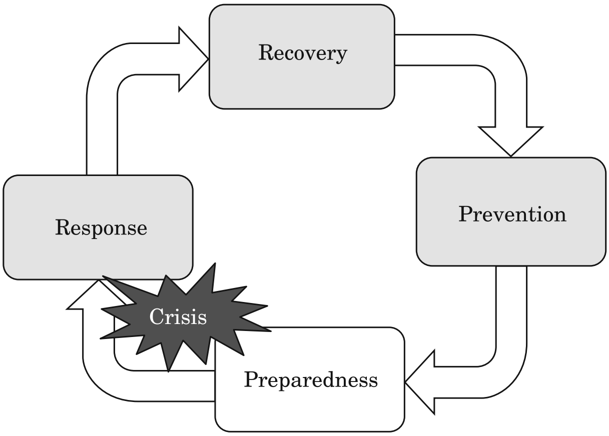


Fig. 2. Crisis management phases
Source: own elaboration based on http://oki.krakow.rzgw.gov.pl.

Prevention phase – the priority in this phase is to take actions, which will eliminate or significantly decrease the probability of a crisis situation occurrence and reduce its consequences. The main objective of undertaken actions is to foil the occurrence of a threat. Should it happen, the prevention phase will focus on minimizing potential effects. Preventive actions mainly deal with the risk analysis, predicting, spatial planning, strategic planning, realisation of investments increasing safety, planning and developing preventive actions, monitoring possible threatening events, systematic classification and identification of a threat sources and evaluation of mental and technical state of the society to the crisis management. However, there are threats which cannot be prevented, for example natural disasters. The only thing that can be done in such situation is to minimise the effects. All actions in this phase are constant, which plays an important role in the whole process of the crisis management. It seems obvious that without properly organised communication the realisation of prevention phase related tasks are impossible or at least very problematic. In this phase, requirements concerning communication revolve around gathering information from multiple sensors, its processing and developing preventive measures.

Preparedness phase – deals with identification of a potential threats, analysis of their character and probability of occurrence. In this phase, it is crucial to formulate roadmaps and operation plans that would be executed during the crisis situations, organise communication as well as warning and alarm systems, establish crisis management centres and provide various services and citizens with proper training of how to behave during the crisis situation. Furthermore, in this phase, actions aiming at increasing manpower and resources necessary for effective reactions are taken. One should remember that communication plans (usually formulated in several versions – for instance in the event of damage of a part of communication infrastructure, among others) shall be developed in this phase.

Response phase – is about taking actions initiating rescue services work, directly responsible for eliminating or limiting the scope of arising crisis situations, which should provide victims with necessary help and neutralize sources of a threat. It's a phase of practical actions that are taken in the event of the crisis situation. Without compelling and well-organised communication those actions cannot be taken efficiently. Properly organised communication is a key to rescue actions. Constantly informing citizens about current situation state is a vital element in the crisis management. Yet, it could not be generated until efficient information flow is established. Taken actions are determined by the character and the size of an event. At this stage all neglected elements from previous phases can be noticed. Any faults or deficiencies in organisation communication may cause great loses. Reacting relies on scrupulous event analyses, designing set of possible solutions, taking right decisions and coordinating works using means of communication.

Recovery phase – it covers actions, of which main objective is to reconstruct infrastructure and strengthen its former state taking into consideration gained experience so as not to let similar situation happen in the future. Such actions rely on damage assessment, helping citizens, replenishing resources and formulating conclusions. It is important to isolate those elements that have been damaged or destroyed. Taken actions should focus not only on eliminating the effects but also on removing causes determining the occurrence of a given crisis situation. There is a visible interdependence between the crisis management phases as mistakes made in the initial phases generate negative consequences during decision making process. The borders between certain crisis management phases are disappearing, which formulate decision making to a continuous process.

From the communication system point of view, the most challenging is the response phase. The response phase is characterized by a variety of emergency plans that are put into action in order to rescue lives and minimize the damage in a disaster or emergency situation. Response is the phase where actual rescue actions are undertaken; these actions include search and rescue, risk assessment, first responders' actions (Ambulances, Fire brigade, Police, etc). Given the unpredictability of natural disasters, this phase requires that first responders conduct rescue actions in real-time in order to stabilize the situation and avoid further damage. Under these circumstances, the need for exchanging information between participants of the rescue actions is undeniable. Another important principle – from communication system perspective -is the decision making process, as presented in Figure 3.

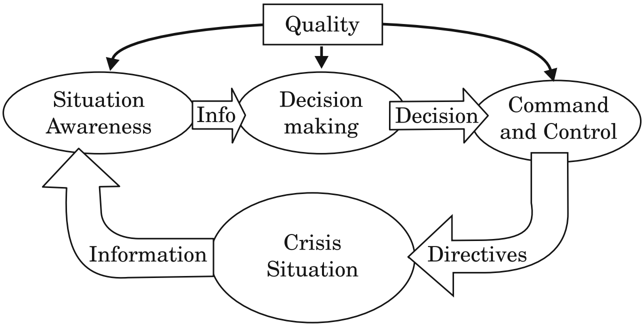


Fig. 3. Crisis management decision making process

The crisis management structure was used in conjunction with the decision-making procedures to show how operational decisions are made. Starting at the Situation Awareness, we can notice that the quality of information available about the situation affects the quality of decision-making, which in turn influences the quality of Command and Control orders and directions. These should have a positive impact on the situation on the ground. New information feeds the tactical and strategic commanders' situational awareness and the circular motion continues. This principle was used as a basis to analyse the capability gap SECRICOM was intended to address.

User requirements in SECRICOM

SECRICOM approach

After reviewing some of the most popular approaches to the requirements engineering, it was decided that none of them fully meets SECRICOM project's needs. As a result, a dedicated approach to the requirement engineering process has been developed within the project. The high-level overview of this approach is presented in the Figure 4.

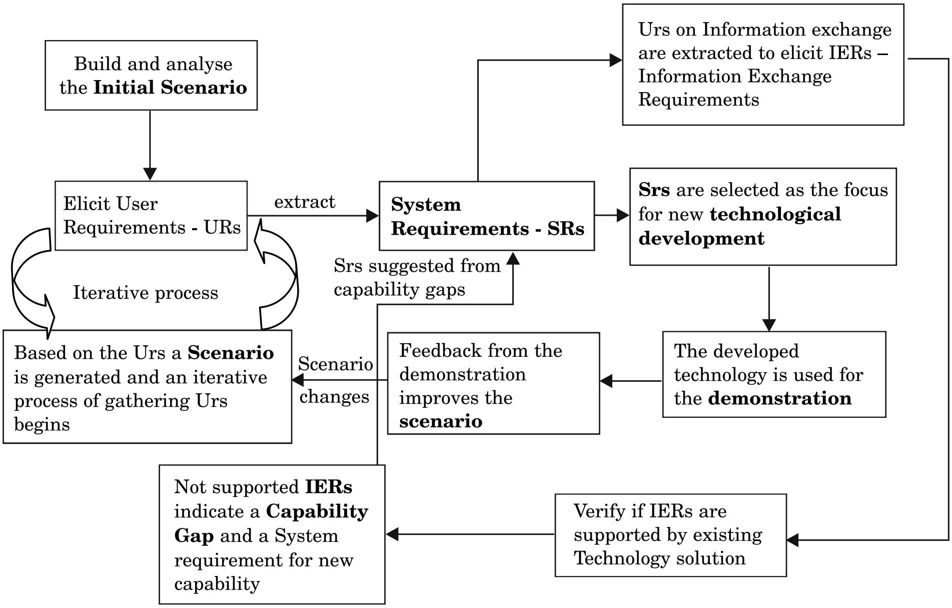

Build and analyse the **Initial Scenario**
Elicit User Requirements - URs
extract
System Requirements - SRs
Urs on Information exchange are extracted to elicit IERs – Information Exchange Requirements
Srs are selected as the focus for new **technological development**
Iterative process
Srs suggested from capability gaps
Based on the Urs a **Scenario** is generated and an iterative process of gathering Urs begins
Scenario changes
Feedback from the demonstration improves the **scenario**
The developed technology is used for the **demonstration**
Not supported **IERs** indicate a **Capability Gap** and a System requirement for new capability
Verify if IERs are supported by existing Technology solution

Fig. 4. SECRICOM Requirements Gathering

Source: own elaboration.

The approach used in SECIRCOM is summarized as follows:

– At the beginning, user requirements are collected from an initial scenario (report or debriefing of an incident). From these URs a new scenario is generated. Afterwards, an iterative process of extracting user requirements that improve the given scenario begins.

– System requirements are selected form the URs, which defines the technology development required.

– Information Exchange Requirements (IERs) are considered the URs that concern exchange of information; IERs are used in gap analysis.

– When an IER is not supported by the existing technology, it is considered a technology capability gap and a new requirement for technology improvements is elicited.

User Focus

Since the SECRICOM project from the very beginning was user-oriented, a User Team had been established for collecting requirements. The team had representatives from various countries as well as agencies to ensure that a complete spectrum of issues is addressed within the project. The B-APCO was responsible for leading this group.

The User Team comprised representatives from Spain, Sweden and UK. These individuals reflected a variety of roles and responsibilities in terms of national, regional and local first responders' agencies. Furthermore, between the members of the user group, they held a range of senior, middle and junior management roles within Fire, Police and Local Authority agencies.

Users constitute the most important source of information in SECRICOM, because it was possible to obtain from them experience based statements, remarks and suggestions from individuals well versed in crisis management and in terms not only of functional requirements, but also non-functional, incl. security requirements for communications assets needed during a major crisis.

IER

In this section Information Exchange Requirements are presented and their use and benefit for defining the requirements for crisis management communication technologies is described.

Information Exchange Requirements (IER) are defined as "the description, in terms of characteristics of the requirement to transfer information between two or more end users. The characteristics described include source, recipients, contents, size, timeliness, security and trigger. IERs are defined to be independent of the communications medium. An IER can express both current and future requirements" (http://ceur-ws.org/Vol-340/paper03.pdf).

Therefore, the purpose of the IERs is to provide structured means for establishing the capability gap and thus defining new requirements. An IER is the *Unconstrained User Requirement for Information Exchange*, thus IERs shall be technology, system and solution independent. IERs are used in capability gap analysis to model URs, which may (or not) be supported by current technology solutions. In order to capture the IERs about an activity or actor, it is necessary to have a well-defined set of URs, and most importantly, these URs should be independent of the technology or system.

To capture the IERs in SECRICOM, the following approach has been applied:

– Define a set of **User Requirements** – describe what users want to achieve.

– A representative **Scenario** brings URs to life.

– The Scenario is broken down into **Activities**.

– Activities are recorded as **IERs**.

– Each activity has one or more IERs associated with it.

– IERs fall into Situational Awareness or Command and Control.

Every operational need for a piece of information to be transferred from one place or person to another has been captured. Then IERs were used in a realistic scenario to ensure that all relevant activities were represented – and in particular to allocate them into Situation Awareness or Command and Control.

IER capture was done in an in-depth exercise during the User Team Exercise in September 2009. In total, over 700 IERs were captured, providing information across a range of criteria including:

– **Source** & **Destination** (of a piece of information).

– **Information Type** (e.g. voice, message, image).

– **Size** (associated with the Information Type).

– **Timeliness** expressed in the worst case as delivery time.

These criteria were complemented with qualitative data on:

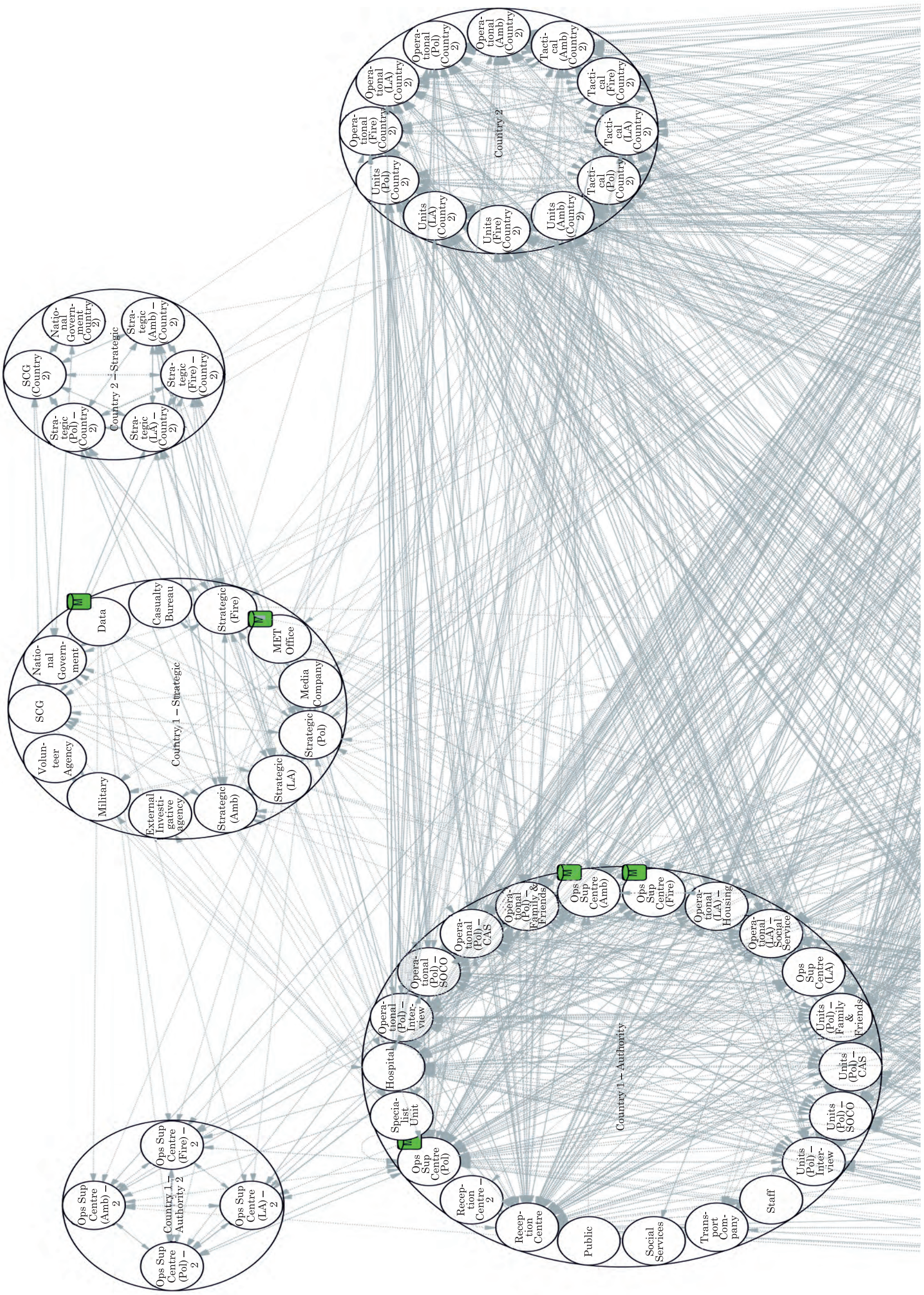
– Criticality.

– Confidentiality.

– Other analysis attributes (e.g. business function).

The complexity of the information exchange is illustrated in Fig. 5, which shows the mapping of IERs to the scenario node location mappings. These location nodes represent a town, a chemical plant, school area, authority areas of different countries, etc. The Fig. 6 presents information flow associated with the scenario and demonstrates co-ordination and liaison with neighbouring country's units and information flows from operational to strategic and fixed infrastructure nodes.

On the other hand, the demand for services – with respect to the agency and command level – have been captured and aggregated. As foreseen, the voice was predominant (for each command level and agency) while the access to the internet was less valuable; those results have confirmed the expectations from interviews. Detailed results are presented in Figure 6.


Country 2
Operational (Pol) (Country 2)
Operational (Amb) (Country 2)
Tactical (Amb) (Country 2)
Tactical (Fire) (Country 2)
Tactical (LA) (Country 2)
Tactical (Pol) (Country 2)
Units (Amb) (Country 2)
Units (Fire) (Country 2)
Units (LA) (Country 2)
Units (Pol) (Country 2)
Operational (Fire) (Country 2)
Operational (LA) (Country 2)
Country 2 – Strategic
National Government (Country 2)
Strategic (Amb) – (Country 2)
Strategic (Fire) – (Country 2)
Strategic (LA) – (Country 2)
Strategic (Pol) – (Country 2)
SCG (Country 2)
Country 1 – Strategic
Data
Casualty Bureau
Strategic (Fire)
MET Office
Media Company
Strategic (Pol)
Strategic (LA)
Strategic (Amb)
External Investigative agency
Military
Volunteer Agency
SCG
National Government
Country 1 – Authority
Operational (Pol) – Family & Friends
Ops Sup Centre (Amb)
Ops Sup Centre (Fire)
Operational (LA) – Housing
Operational (LA) – Social Service
Ops Sup Centre (LA)
Units (Pol) – Family & Friends
Units (Pol) – CAS
Units (Pol) – SOCO
Units (Pol) – Interview
Staff
Transport Company
Social Services
Public
Reception Centre
Reception Centre – 2
Ops Sup Centre (Pol)
Specialist Unit
Hospital
Operational (Pol) – Interview
Operational (Pol) – SOCO
Operational (Pol) – CAS
Country 1 – Authority 2
Ops Sup Centre (Fire) – 2
Ops Sup Centre (LA) – 2
Ops Sup Centre (Pol) – 2
Ops Sup Centre (Amb) – 2

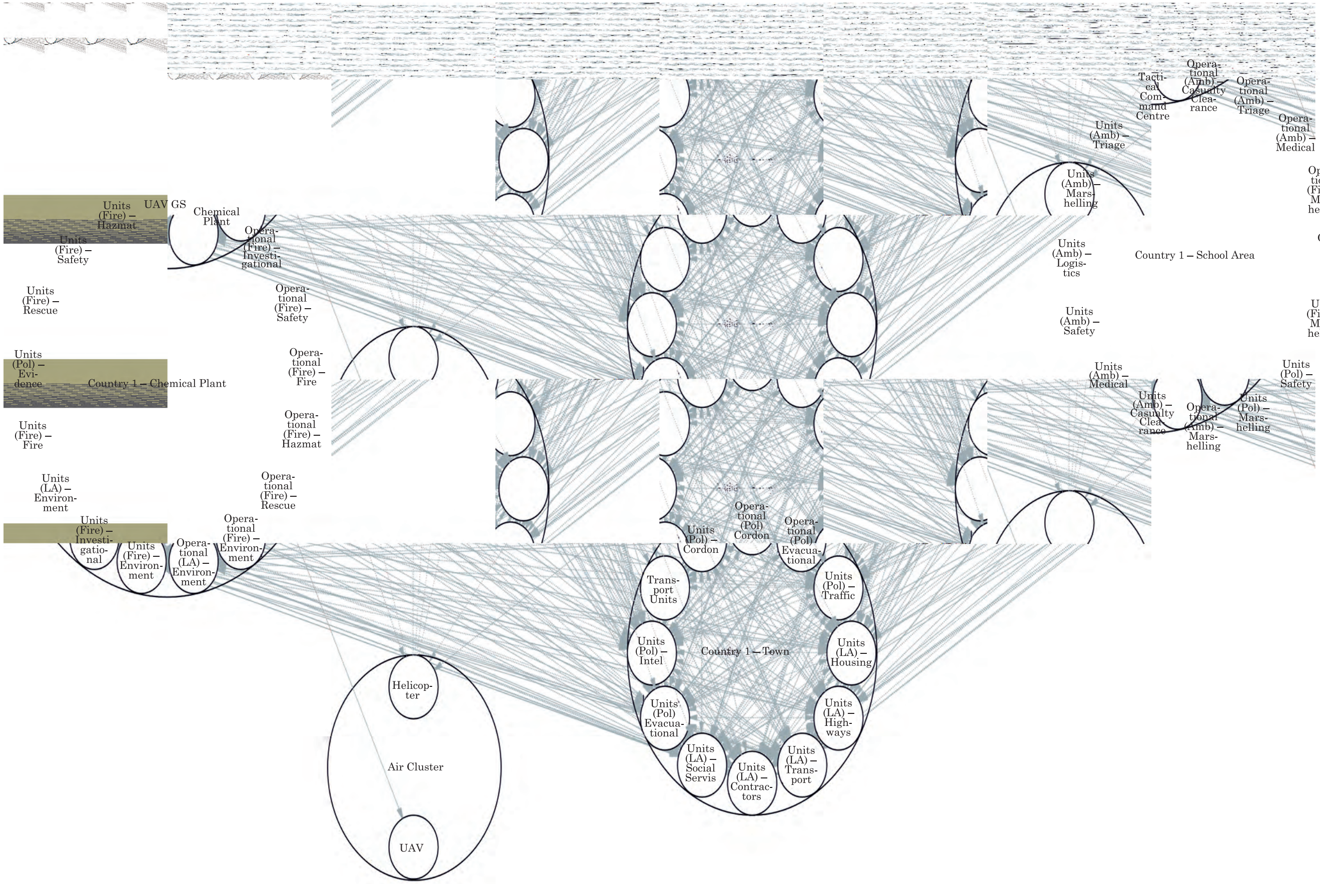

Tactical Command Centre
Operational (Amb) – Casualty Clearance
Operational (Amb) – Triage
Operational (Amb) – Medical
Units (Amb) – Triage
Units (Amb) – Marshelling
Operational (Fire) – Marshelling
Operational
Units (Amb) – Logistics
Country 1 – School Area
Units (Amb) – Safety
Units (Fire) – Marshelling
Units (Amb) – Medical
Units (Pol) – Safety
Units (Amb) – Casualty Clearance
Operational (Amb) – Marshelling
Units (Pol) – Marshelling
Units (Fire) – Hazmat
UAV GS
Chemical Plant
Operational (Fire) – Investigational
Units (Fire) – Safety
Units (Fire) – Rescue
Operational (Fire) – Safety
Operational (Fire) – Fire
Units (Pol) – Evidence
Country 1 – Chemical Plant
Units (Fire) – Fire
Operational (Fire) – Hazmat
Units (LA) – Environment
Operational (Fire) – Rescue
Operational (Fire) – Environment
Units (Fire) – Investigational
Units (Fire) – Environment
Operational (LA) – Environment
Units (Pol) – Cordon
Operational (Pol) Cordon
Operational (Pol) Evacuational
Transport Units
Units (Pol) – Traffic
Units (Pol) – Intel
Country 1 – Town
Units (LA) – Housing
Units (Pol) Evacuational
Units (LA) – Highways
Units (LA) – Social Servis
Units (LA) – Contractors
Units (LA) – Transport
Helicopter
Air Cluster
UAV

Fig. 5. Information flow between agencies

Source: SECRICOM Information Exchange Requirements 1.0.

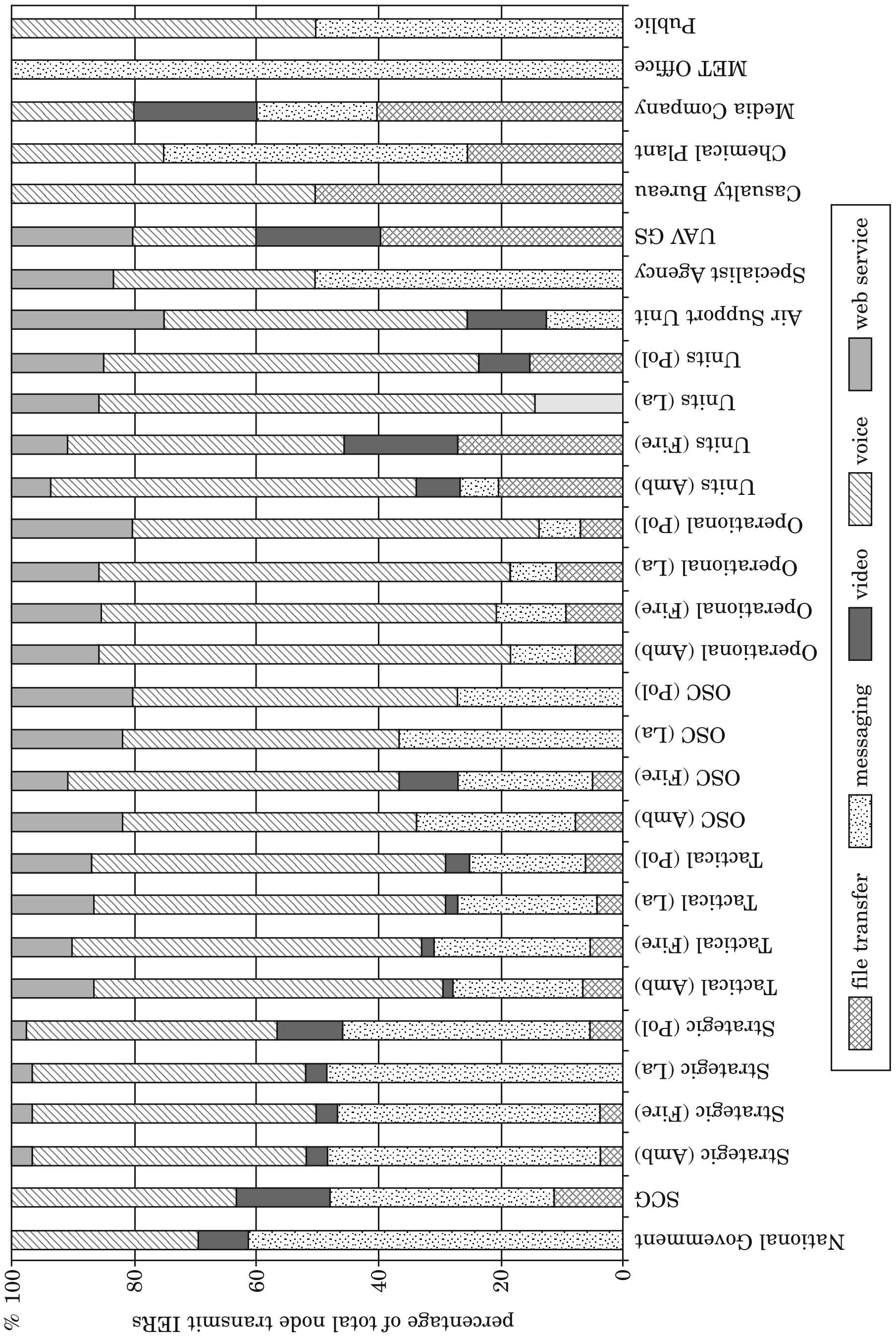


Fig. 6. Total Node transmitting IERs

Source: SECRICOM Information Exchange Requirements 1.0.

During the next step, the IERs were **developed** from the URs **analysed** in the context of a scenario, and used to **model** existing communications architecture and to **identify** which IERs would be supported by the current architecture. Any **unsupported IERs** would tend to indicate a Capability – and therefore an Interoperability – shortfall (see Fig. 7 below).

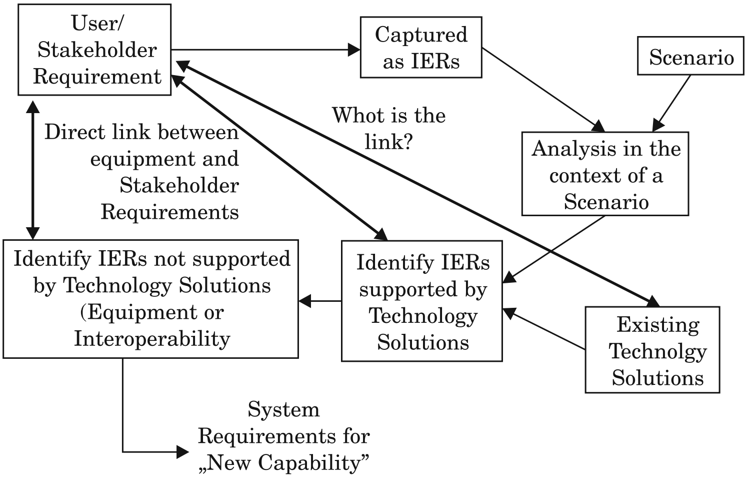

User/ Stakeholder Requirement
Captured as IERs
Scenario
Whot is the link?
Direct link between equipment and Stakeholder Requirements
Analysis in the context of a Scenario
Identify IERs not supported by Technology Solutions (Equipment or Interoperability
Identify IERs supported by Technology Solutions
Existing Technolgy Solutions
System Requirements for „New Capability"

Fig. 7. Capability gap analysis process
Source: SECRICOM Information Exchange Requirements 1.0.

Gap analysis

The purpose of the gap analysis in the SECRICOM project was to find the problems of existing communication technologies when used in crisis situations, and mitigate key capability gaps faced by users of existing systems. The SECRICOM project addressed not only existing capabilities but also new ones. SECRICOM provides value added through finding new requirements for future technology developments. This is achieved at the overlap between existing and future infrastructure and systems, as illustrated in Figure 8.

To perform gap analysis a realistic operational scenario was used, user requirements were formulated and then IERs were extracted from URs.

User requirements were extracted from the following sources:

- Interviews with end-users.
- Scenario-based demonstrations.

IERs provided the communications requirements upon which the technical development would be based – and which would also direct the final demonstration.

The idea was to identify URs that are not met by existing communication systems and mark it as a capability gap. In this way, the capability gaps provide the value added by SECRICOM as they are interpreted as system requirements for technological development.

Additionally, in the SECRICOM the capability gap approach was used as a basis to develop URs and was verified in a demonstration test.

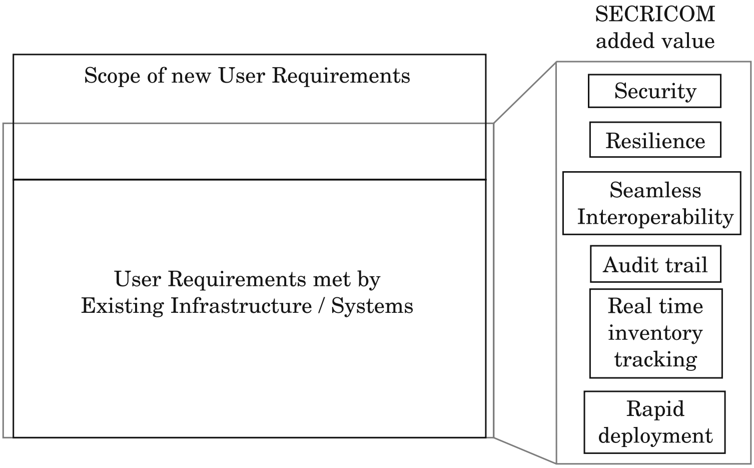

SECRICOM added value
Scope of new User Requirements
User Requirements met by Existing Infrastructure / Systems
Security
Resilience
Seamless Interoperability
Audit trail
Real time inventory tracking
Rapid deployment

Fig. 8. SECRICOM added value

Source: SECRICOM project.

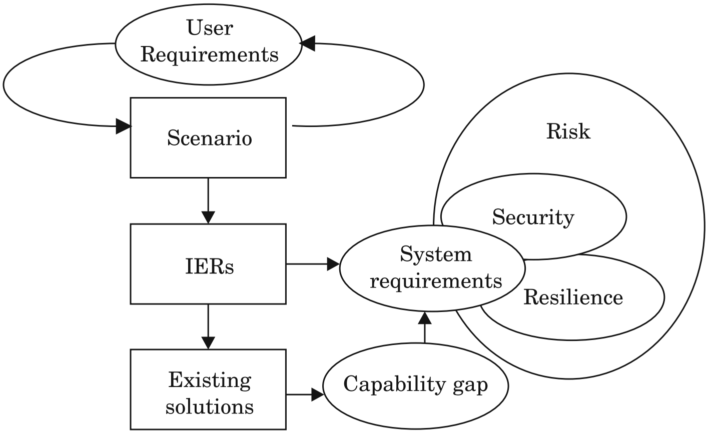


Fig. 9. Requirements within the SECRICOM project

Source: Security requirements user workshop – Report.

In order to demonstrate the significance of the capability gap outcome. A comparison between the SECRICOM User Requirements and current communications systems (e.g.: TETRA, TETRAPOL, WiMAX, CB, GSM/UMTS) capabilities needed to be made. This comparison established the difference / delta between where the capability shortfall exists and where SECRICOM could add value. As it was explained previously, in the SECRICOM project user requirements were created from the scenario through an iterative process. Those URs, on the other hand, were feeding the system requirements. Selection of the appropriated system requirements defined the technology development required. A vignette or mini-scenario, together with the new technology capability, formed the final demonstration. Figure 9 presents the methodology used for acquiring system requirements.

Security requirements

Another aspect that has been discussed through the analysis is security. Security was always an integral part of the project. To correctly address the security requirements, a workshop – dedicated just to security issues – with the user team was conducted. The requirements that emerged are framed around a crisis management structure; the security requirements vary on each of the three levels of command chain. Furthermore, the impact analysis was performed; hence the operational capabilities related with each asset have been established, referring to:

– Saving lives.
– Ability to conduct emergency services.
– Provision of local contingency services.
– Impact on judicial proceedings.
– Impact on foreign relations.

Additionally the security requirements in terms of confidentiality, integrity and availability have been established for the most important services, namely:

– File transfer (documents).
– Messaging (Email, Data/Image).
– Video.
– Voice.
– Access to the Internet.

The outcome of the security assessment was understanding of how important the identified factors are to the users. Bellow the main results achieved by the security assessment are summarised:

– Voice communications at all 3 levels of command, and between agencies, are seen as critical. It requires the **highest level of security** in terms of Confidentiality, Integrity and Availability.

– Messages and file transfer are seen as the **second important**.

– Access to the Internet is the **least valued**.

– Integrity, across all 3 command levels, is seen as a **key requirement** (voice in particular) for all communications assets.

In comparison to Integrity and Availability, Confidentiality is considered a **less important requirement**.

– Availability.

– Voice viewed as **essential**.

– Messaging and file transfer more important than video and web.

The following charts (Fig. 10, Fig. 11, Fig. 12) illustrate the results at strategic, tactical and operational level considering the impact perspective on confidentiality, integrity and availability, for the most dominant services.

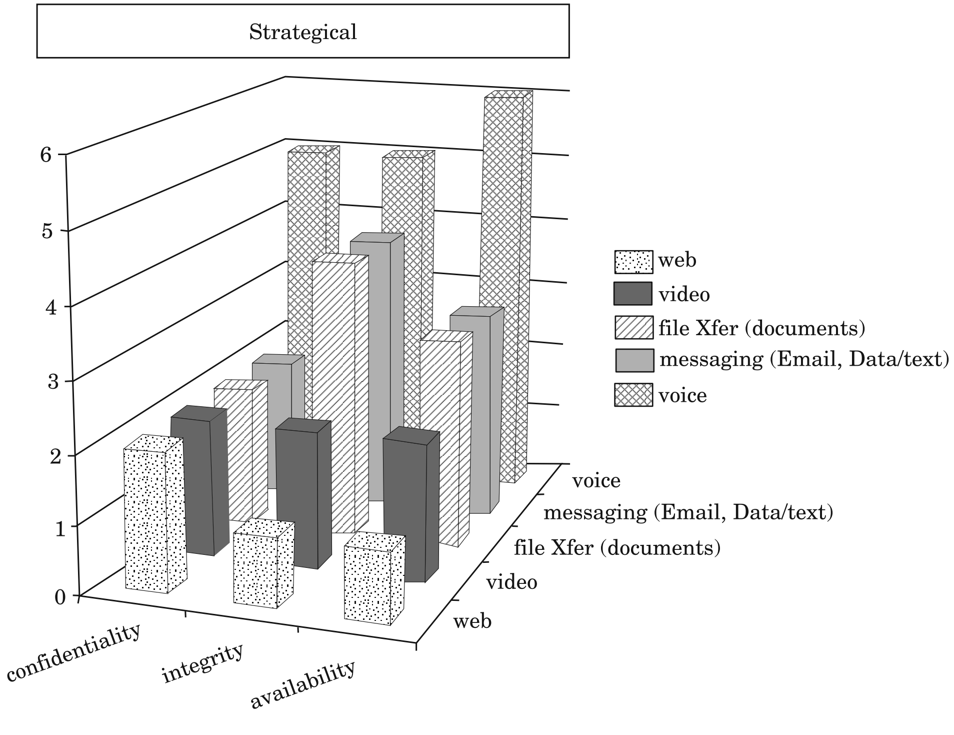


Fig. 10. Security requirements at strategic level
Source: Security requirements user workshop – Report.

The key findings of the IER exercised as a whole are:

– Overall voice is predominant (~ 50%), messaging is next (~ 25%).

– Voice more concentrated at operational level – decreases higher up the command chain.

– Data more concentrated at strategic level – decreases lower down the command chain.

– Specific increased need was identified for image and video capabilities at operational level.

– **Intra**-Agency communications are key at all levels of command.

– **Inter**-Agency communications account for nearly a quarter of all IERs.

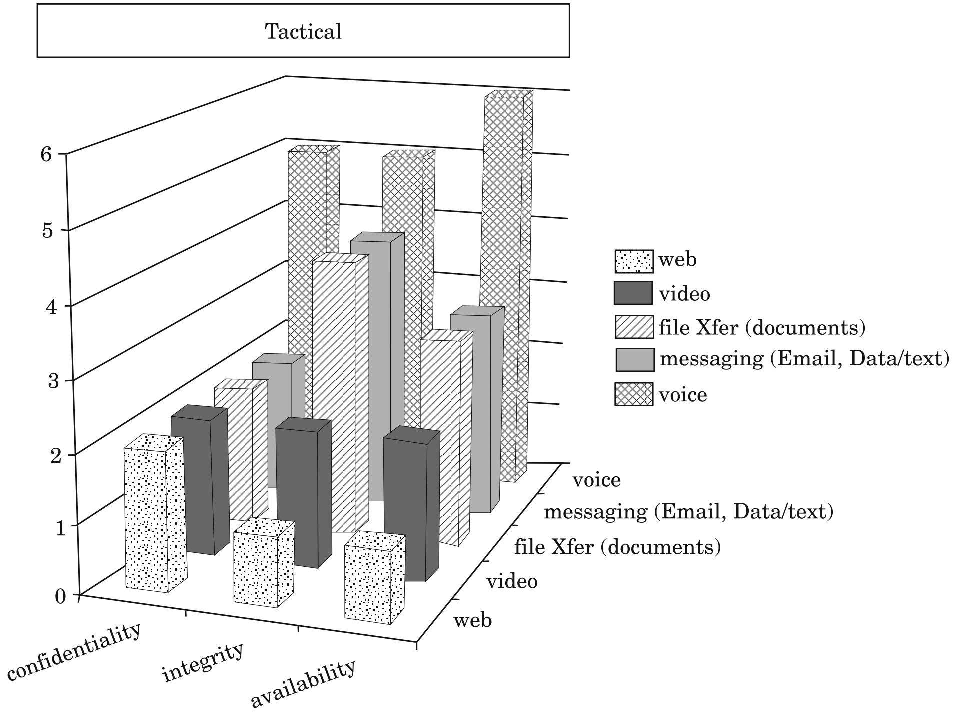


Fig. 11. Security requirements at tactical level
Source: Security requirements user workshop – Report.

– Situational Awareness provides the greatest proportion of IERs (~ 59%).

– Ratio of Command & Control to Situational Awareness distorted due to voice.

– Data versions of the same IER (driven by the need of audit trail).

– Voice remains the most significant IER data type for both Command & Control and Situational Awareness (Situational Awareness demands a greater use of non-voice data types).

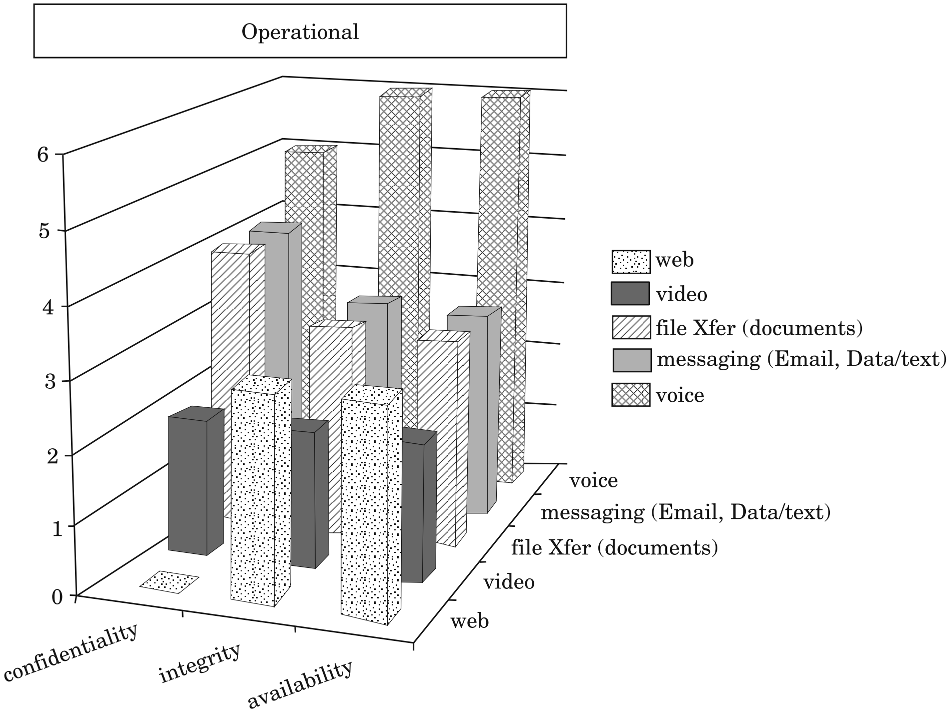


Fig. 12. Security requirements at operational level
Source: Security requirements user workshop – Report.

Conclusions

In this article an approach to elicitation, analysis and validation of user requirements for critical system has been presented. A brief review through most popular approaches to the user requirements have been given, taking into account its use in the SECRICOM project. The system environment (incl. structure, principles and decision making process in crisis management) was analysed and results were reflected in the proposed process.

Afterwards a detailed description of SECRICOM requirements gathering process has been provided. This process is based on three complementary approaches:

- Permanent monitoring of the technical innovations in terms of their adaptation in the area of crisis management, with the particular focus on the communication area.
- Use of scenarios – historical knowledge to develop two types of scenarios that verify the effectiveness of particular features:
  - Generic scenarios,

– Anticipation scenarios (which assume a range of possible outcomes).
– Gap analysis.

This approach combines both retrospective and prediction elements. On top of that, it allows to monitor and proper address important changes in communication domain.

The most important change in crisis management communication is a growing use of ICT systems. Although symptoms of these changes were evident for many years, the current state of telecommunications infrastructure allows to offer a wide range of new services, such as analysis of connections in real-time, video telephony, use of new source of information (e.g. social networks) or dynamic group management.

Finally, the requirements obtained were presented. The results were presented for each command chain level. There was a particular focus on security issues, since they were found critical in crisis management field.

Acknowledgments

This work was funded by the SECRICOM project (EC FP7-SEC-2007 grant 218123).

Translated by AUTHORS

Accepted for print 30.06.2012

References

ABRAN A. 2004. *Guide to the Software Engineering Body of Knowledge*. IEEE Computer Society.
Advertising materials http://www.itti.com.pl/pl/projekty-ue/projekty-trwajace2.html.
SECRICOM D2.2 Crisis management requirements assessment report.
SECRICOM D9.1.1 requirements of the security model.
GOBAN-KLAS T., SIENKIEWICZ P. 1999. *Społeczeństwo informacyjne: Szanse, zagrożenia, wyzwania.* Wydawnictwo Postępu Telekomunikacji, Kraków.
http://ceur-ws.org/Vol-340/paper03.pdf
http://int3.de/res/RUP/RUP_Paper_JohannesPassing.pdf
http://spot.pcc.edu/~lmiddlet/CIS122/coursematerial/Requirements.html
http://www.comp.lancs.ac.uk/computing/resources/IanS/SE7/index.html
HULICKI Z. 1998. *Systemy komunikacji multimedialnej*. Wydawnictwo Postępu Telekomunikacji, Kraków.
Information from the website http://www.secricom.eu/
KOTONYA G., SOMMERVILLE I. 1998. *Requirements Engineering: Processes and Techniques Chichester.* UK: John Wiley and Sons.
Risk management Guide for Information Technology Systems.
SECRICOM Information Exchange Requirements 1.0. R. Edwards, J. Dexter, S. O'Neill – Report, internal SECRICOM deliverable.
Security requirements user workshop – Report, internal SECRICOM deliverable.
WOJCIECHOWICZ W., ZYCH J. 2010. *Kierownicze gry decyzyjne w zarządzaniu kryzysowym*. MAIUSCULA, Poznań.
WOJCIECHOWICZ W., ZYCH J. 2011. *Koncepcja infrastruktury telekomunikacyjnej o podwyższonej niezawodności*. In: *Bezpieczeństwo współczesnego świata – Informatyka, technika i gospodarka*, Ed. Z. Dziemianko, WSHiU, Poznań.

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

INFORMATION AND COMMUNICATION TECHNOLOGY AND CRISIS MANAGEMENT

***Wojciech Wojciechowicz*[1,2], *Jan Zych*[2], *Witold Hołubowicz*[2,3]**

[1] Institute of Computing Science, Poznań University of Technology
[2] ITTI sp. z o.o.
[3] Adam Mickiewicz University of Poznań

Key words: crisis management, ICT, communication.

Abstract

In the present article selected telecommunication aspects in the area of crisis management are exposed. In particular, the focus is put on the response phase, since there are new challenges at the junction of organisational and technical layers, incl. interoperability, new functionalities and models. Those aspects have not been exhaustively tested in real situations; thus such issues still require multiple testing, verification and validation. In this article the communication problems are collated with new solutions, such as the use of cloud computing, social media and additional functionalities to increase the security level. The main aim of this article is to introduce challenges, as well as new opportunities provided by the implementation of new Information and Communication Technologies in the area of crisis management.

ZAGADNIENIA TELEINFORMATYCZNE A ZARZĄDZANIE KRYZYSOWE

***Wojciech Wojciechowicz*[1,2], *Jan Zych*[2], *Witold Hołubowicz*[2,3]**

[1] Instytut Informatyki, Politechnika Poznańska
[2] ITTI sp. z o.o.
[3] Uniwersytet im. Adama Mickiewicza w Poznaniu

Słowa kluczowe: zarządzanie kryzysowe, ICT, łączność.

Abstrakt

W artykule opisano wybrane problemy telekomunikacyjne, które identyfikuje się w obszarze zarządzania kryzysowego. Szczególny nacisk położono na fazę reagowania. Na styku warstwy organizacyjnej i technicznej pojawiają się nowe problemy, np. z interoperacyjnością, nowymi funkcjonalnościami technicznymi, które nie zostały jeszcze dostatecznie sprawdzone w praktyce, a to wymaga wielowątkowych badań. Autorzy diagnozują problemy z obszaru łączności, uwzględniając takie zagadnienia, jak: zastosowanie chmury obliczeniowej, wykorzystanie sieci społecznościowych czy implementacji nowych funkcjonalności na podstawie istniejącej infrastruktury telekomunikacyjnej. Głównym celem artykułu jest przedstawienie zarówno wyzwań, jak i szans dla zarządzania kryzysowego, płynących z nowych rozwiązań teleinformatycznych.

Introduction

In recent years, there has been a significant change in the perception of challenges in the field of crisis management (CM). Instead of focusing on a major incident, there is a trend to draw special attention to several smaller incidents, which overlap/accumulate in a relatively short period of time. A perfect example illustrating this were the national crisis management system exercises – LIBERO II, organised by the Government Security Centre in collaboration with the Ministry of Internal Affairs and the PL.2012 company. During the event many scenarios were practiced, incl. energy failure, a disaster on an expressway, removing the effects of civil unrest after the football match. The scope of the exercises did not refer to one big event, but to many events that correlated with each other.

Furthermore, there is an increasing awareness of other crisis management aspects, such as use of modern information and communication technology, and the search for communication platforms able to provide adequate functionalities. News from social networks is of great importance, and is exceptionally useful during decision making process. It turns out, that the first thing that accidental witnesses tend to do is to post multimedia material on social networks, instead of reporting it to the appropriate crisis management agencies. Thus, there is a need for crisis management agencies to permanently monitor those social networks with respect to extraction of useful, up to date and unique pieces of information about given incident. In certain cases, it may become the only factual evidence.

The main scientific problem was to identify communication challenges in the field of crisis management, and to propose new solutions in this field. In the article the focus was put on broad context of crisis management rather than on a single incident.

Challenges in crisis management

Crisis management involves the coordination of activities by different groups in an effort to avoid or minimize disaster impact. This section describes the challenges that need to be confronted by crisis management in order to be more effective. There are four phases in crisis management:

- prevention,
- preparedness,
- response,
- recovery (PIĄTEK 2006).

Where the most challenging – from the crisis management communication system perspective – is the response phase. That is the area where rescue actions are executed; given the specificity of the domain, this phase is the most dynamic. It is very difficult to precisely plan the response in advance, since it is not possible to predict when and where crises will occur. In fact, response actions always need to be undertaken in real time. Moreover, disasters often affect communication systems (e.g. natural disasters like floods or fires can seriously harm infrastructure). Therefore, the least that can be done is to prepare for the effects.

When it comes to ensuring communication between all stakeholders responsible for crisis management it is important to identify threats in real time, and coordinate the rescue actions accordingly to them. Participation of the Police, National Fire Service and Emergency Medical Service in training exercises confirmed that when an incident occurs, the number of connections to the crisis management centres increases significantly. It means that requests with the same content but through different channels (text, voice, video) are made. Quick extraction of the most important part from each message, its processing and passing to the decision makers, responsible for coordinating rescue actions in the crisis management centres, is extremely challenging, both at technical and organisational levels.

On top of that, multiple agencies as well as volunteers and passers-by are involved in the actions at the crisis scene. This creates yet another challenge which is the need to ensure communication and effective flow of information between them. That, combined with the response phase characteristics, puts an additional pressure on communication systems.

Furthermore, crisis management involves the execution of very broad and diverse actions. Possible incidents include:

- riots during mass event,
- natural disasters (earthquake, avalanche, flood, wildfire),
- man-made/terrorist attack/confrontations,
- workplace violence/misdeeds,
- accidents (crashes on land, sea and air).

For those incidents – each supporting by technology – a number of challenges might be identified. In high level of abstraction, they can be categorised into technical and organisational ones, often influencing each other. The key technical challenges include (*Metody sztucznej inteligencji.* 2009, ŚWIĄTNICKI, ŚWIĄTNICKI 1992):

- insufficient spectrum of services,
- lack of interoperability between various ICT systems,
- security aspects,
- managing user groups in real-time,
- effective information dissemination (incl. notifications to the population),

while the organisational:

- units autonomy,

– frequency allocation,

– cooperation with telecommunication operators,

– improving the work of the international community in several areas, such as decision making and situational awareness.

First responders have invested in many – often incompatible – telecommunication solutions. As a result, the possibility of exchanging information in an effective way has been reduced. In the next sections we will discuss the aforementioned technical challenges.

The technical challenges in the field of crisis management

This section provides a short description of technical challenges in the crisis management field. The current status of technical challenges is presented first, then a short description of future perspectives is given; and at the end we provide examples of how difficult crises are becoming with the purpose of illustrating the need for confronting the technical challenges.

State of the art

As it was stated previously, the response phase puts an extra pressure on the telecommunication systems. Unpredictability of time and place results in urgent needs for extra capabilities at the incident scene.

On top of that, agencies and some command chain levels have different user requirements – not only functional requirements which may concern different spectrum of services available but also non-functional (like security or performance). Additionally, the requirements may vary depending on the incident type.

Recently some investments have been done in various technologies supporting first responders. However, these investments have been done without taking into consideration the exchange of information between different crisis management actors; hence there is no unification of infrastructure, end-user devices and software applications. As a result, full interoperability is not achievable.

Perspectives

Currently, there is no single vision for the crisis management telecommunication system. Many agencies have done some technology updates without obtaining the consent of other stakeholders. There is also a lack of a *perfect* telecommunication solution, which will meet the needs of each stakeholder (in terms of characteristics and costs). Furthermore, such a solution have not been designed yet.

Terrestrial Trunked Radio (TETRA) is an open standard, created by ETSI for dispatching digital radio-telephone communication [(Terrestrial Trunked Radio (TETRA)a]. It could be said that TETRA is the optimal choice; but its efficiency is still doubtful. One may notice that this solution is getting more and more "mature", and not even TEDS [(Terrestrial Trunked Radio (TETRA)b] (TETRA Enhanced Data Services) may guarantee the sufficient bandwidth when multimedia services are in use. Also the costs (both implementation and maintenance costs) are believed to be drawbacks of TETRA.

Crisis globalisation

The larger and larger crises, the greater number of agencies need to be involved, which puts an extra pressure on the crisis management communication system. The incidents (esp. natural ones, like floods and wildfire) also occur in cross-border areas, which require greater cooperation between nations as rescue actions involve rescue teams from several countries (vide Central European flooding in 1997 and 2010 or Haiti earthquake and Katrina hurricane in USA).There are also other types of incidents, such as terrorist attacks (incl. Madrid train bombings on 11 March, 2004, 2011 Norway attacks, 7/7 London bombings) or cyber threats (incl. attacks on SCADA systems) that have already become crises.

The crises also vary in terms of specific telecommunication needs and background, e.g.

– Prevention of disorders (e.g. riots during mass events) – a large number of volunteers are engaged, as well as third-party organisations (incl. personal security). Due to law requirements (e.g. banned access to the event for selected persons) the access to external databases would be appreciated.

– Flood – mostly involves several agencies – and also third-party organisations (e.g. construction companies) as well as volunteers, but it is almost impossible to indicate them in advance. Thus, there is a need for quick and seamless inclusion of new actors into the system. Another very important feature is the notification of residents. That could reduce disinformation as well as improve the coordination at the scene.

– Terrorist attack – where behaviour detection would have great importance. Audit trial could be very useful for public prosecutors. Different security requirements could be defined, since e.g. confidentiality may play a great role. Also actor localisation could be useful not only in terms of coordinating the rescue teams, but also rescuing the victims (using e.g. their mobile phones) (*Bezpieczeństwo...* 2009).

The importance of ICT in crisis management is undeniable; communication is necessary in order to achieve effective coordination. ICT provides efficient coordination of activities, the sharing of information between organisations working at the scene as well as access to new data and databases, such as images, maps, and infrastructure information. The role of ICT is getting broader, since not only new threats are identified, but also new opportunities are visualised.

New opportunities for crisis management

Nowadays, crisis management is not able to operate efficiently without the support of the state of the art ICT. In order to find an optimal operation model in crisis management it becomes more frequent to take the advantage of various technological innovations (e.g. trusted computing and agent-based infrastructure) or organisational solutions (e.g. cloud computing). In this section, an outline of new technological opportunities for improving crisis management is provided.

Mobile technologies

Currently, mobile technology is advancing rapidly, both in terms of mobile phone popularity and capabilities. Modern mobile devices (palmtops, mobile phones, etc.) are capable of performing tasks that used to be reserved for personal computers.

With regard to capabilities, there is a marked trend to integrate hitherto separate devices into a single solution. Modern mobile devices are often equipped with auto-focus, a digital camera with several Mega pixels, Full HD video recording possibility (such resolution was barely achievable for dedicated digital cameras just a couple of years ago); moreover, these devices have several GBs of internal storage (with possibility to further increase using flash memory). Combined with broadband (e.g. based on HSDPA or WiFi b/g/n) data transmission and access to modern services (e.g.: online maps (even with traffic information and predictions), weather forecast or social media) mobile phones are considered to be a great tool in crisis management, used not only for communication between responders, but also for dissemination of information among the public in general (FICOŃ 2007).

Social media

Social media is set of technologies that allow people to exchange multimedia information. Despite the fact that the information in social media comes from sources that are not verified, social media allow people to exchange information, ideas, opinions and experience. Therefore, social media has become very popular and this trend is growing.

The example of the 2010 flooding in Central Europe emphasizes the importance of using social media during crises situations. Citizens of the Bydgoszcz city in Poland were using a forum to inform each other about the water level; this source of information was much more effective than official communiques in traditional media. Nevertheless, with the growing popularity of social media, this information could be disseminated even faster -using applications like Facebook, Twitter, Web log and others. Comparing social media to internet forums, one may notice that the former one allows to exchange information almost in real time e.g. through smartphones. The cost in terms of development and maintenance of infrastructure as well as disseminating the information to many recipients is negligible, since social media does not require any additional costs apart from the Internet connection bills.

Cloud computing

Another new opportunity for crisis management is the use of dedicated services in modern business models – cloud computing. The main idea behind cloud computing is to provide services from remote centres using the Internet as a communication channel. In other words, cloud computing provides applications that run on the Internet. Cloud computer services are divided into four models, according to the capability provided (VOORSLUYS et al. 2011):

– IaaS – Infrastructure as a Service. This model provides all the equipment needed by an organisation to support operations, it includes hardware, servers, storage and network components. In this model, the cloud provider is responsible for maintaining the equipment.

– PaaS – Platform as a Service. In this model, cloud providers deliver a computing platform including an operating system, a programming language execution environment, database and web server. With PaaS applications developers can design, run and debug their software solutions on a cloud platform, and do not have to worry about buying and maintaining the hardware and software layers.

– SaaS – Software as a Service. It comprises software applications that are installed on the cloud and that can be accessed by cloud users. Since the software applications are located on central hosts, the cloud users can access them through a browser. In SaaS, users do not have to maintain the data and infrastructure on which the application is running e.g. games, google docs, e-mail, etc.

– BPaaS – Business Process as a Service. This model includes any business processes delivered as a service over the Internet (for example, payroll, printing, e-commerce) and accessible by multiple web-enabled interfaces and devices such as PC, tablets and smartphones.

Cloud computing could contribute to crisis management by facilitating information sharing among first responders at different management levels (central, regional and local), and making the emergency notification more accessible to the public. Additionally, cloud computing reduces costs when it comes to data storage and recovery after a disaster. Companies that own the infrastructure locally could be severely affected by a disaster as their server may be permanently destroyed and backup may be lost. In the case of a disaster affecting a cloud computing data centre, user data will not be lost since suppliers of cloud infrastructure replicate user data and cloud servers across multiple data centres.

Furthermore, the data stored on the cloud is highly secured by cloud providers. In the data centres, the integrity of the information is protected with power generators, monitoring systems and 24/7 security personnel as well as technical specialists.

There is a wide range of possible cloud computing applications in crisis management. It not only improves the current services (in terms of e.g. costs, scalability, confidentiality, availability, security, redundancy and performance), but also provides new opportunities.

Commercial ICT in crisis management

In the previous section selected new opportunities for crisis management have been presented. Mostly they are related to the use of commercial ICT equipment in the field of public safety. Even if the infrastructure is designed for commercial use, it still can bring significant value to the crisis management field.

Current telecommunication infrastructures (e.g.: WiMAX, 3G, LTE or even WiFi) enables a wide range of services desirable in the crisis management field, like:

- (broadband) IP transmission,
- high capabilities:
  - many simultaneous voice calls,

  - video calls,
  - broadband data transmission,
  - low latency,
  - low error-rate,
- redundancy,
- positioning services,
- access to external data sources,
- a wide range of compatible, off-the-shelf devices,
- great network coverage (almost 100% in EU),
- access to numerous external services, incl.:
  - social networks,
  - on-line maps,
  - current traffic information and forecast,
  - weather forecasts (Ficoń 2007).

At the same time, it provides an easy (often cost-effective) possibility to implement additional services:

- increased encryption,
- audit trial,
- information broadcast,
- online group management.

Further actions and conclusions

In this article selected challenges as well as opportunities in crisis management have been presented. The authors have underlined the possibility to use commercial infrastructure – with respective risks and chances – in the crisis management field.

The aim of this article has been achieved by identification of new functionalities for communication systems in crisis management, at the junction of the organisational and technical layers. The research is presented in a general context, with the focus on a number of innovations that can be adopted in the area of crisis management (e.g., social media, cloud computing, mobile phones). Current results are proofs of concepts rather than ready-to-use solutions. In addition, various approaches have been undertaken, incl. short, medium and long-term solutions, but no common vision has been established.

It is worth mentioning that no ultimate telecommunication solution for crisis management is available or expected to appear soon. As for today, the research on this problem has been undertaken by several initiatives – incl. research FP7 projects like:

- SAFECOM,

– EULER – EUropean Software Defined radio for wireless in joint security operations,

– MESA – Mobile Broadband for Public Safety,

– HIT-GATE – HIT-GATE – Heterogeneous Interoperable Transportable GATEway for First-Responders,

– SECRICOM – Seamless Communication for Crisis Management for EU safety,

– FREESIC – Free Secure Interoperable Communications.

However, the problem is still open.

Acknowledgments

This work was funded by the SECRICOM project (EC FP7-SEC 2007 grant 218123).

Translated by AUTHORS

Accepted for print 30.06.2012

References

Bezpieczeństwo w środowisku lokalnym. 2009. Red. W. Fehler, Arte, Warszawa.

FICOŃ K. 2007. *Inżynieria zarządzania kryzysowego. Podejście systemowe*. Bel Studio, Warszawa.

http://kbn.icm.edu.pl/gsi/raport.html

http://tetraforum.pl/10-lat-tetry-w-polsce.html

http://www.alert-sms.pl/alert-samorzadowy.php

http://www.ipedr.com/vol25/25-ICEME2011-N00035.pdf

http://www.sisms.pl/pl/glowna/ostrzeganie-przed-zagrozeniami.html

LEVINSON P. 2006. *Telefon komórkowy. Jak zmienił świat najbardziej mobilny ze środków komunikacji*. Muza, Warszawa.

LIDWA W. 2010. *Zarządzanie w sytuacjach kryzysowych*. Akademia Obrony Narodowej, Warszawa.

Metody sztucznej inteligencji. 2009. Eds. E. Nawarecki, G. Dobrowolski, M. Kisiel-Dorohinicki, Wydawnictwo AGH, Kraków, p. 19–22, 219–253.

PIĄTEK Z. 2006. *Procedury i przedsięwzięcia systemu reagowania*. Akademia Obrony Narodowej, Warszawa.

SIENKIEWICZ K. 2010. *Zarządzanie kryzysowe w administracji publicznej*. Difin SA, Warszawa.

ŚWIĄTNICKI W., ŚWIĄTNICKI Z. 1992. *Bronie inteligentne*. Wydawnictwo Bellona, Warszawa, p. 8, 9.

Terrestrial Trunked Radio (TETRA)a; Release 2; Designer's Guide; TETRA High-Speed Data (HSD); TETRA Enhanced Data Service (TEDS).

Terrestrial Trunked Radio (TETRA)b; Voice plus Data (V+D); Part 17: TETRA V+D and DMO specifications; Sub-part 4: Release 2.0.

VOORSLUYS W., BROBERG J., BUYYA R. 2011. *Introduction to Cloud Computing*. In: *Cloud Computing: Principles and Paradigms*. Eds. R. Buyya, J. Broberg, A. Goscinski. New York, USA: Wiley Press. pp. 1–44. ISBN 978-0-470-88799-8. http://media.johnwiley.com.au/product-data/excerpt/90/04708879/0470887990-180.pdf.

ZYCH J. 2010. *Metodologia badań bezpieczeństwa narodowego*. In: *Metody badawcze w obszarze bezpieczeństwa narodowego*. Eds. P. Sienkiewicz, M. Marszałek, H. Świeboda, AON, Warszawa.

ZYCH J. 2010. *Nowe wyzwania i wykorzystanie współczesnej nauki w zarządzaniu kryzysowym*. In: *Gry decyzyjne w zarządzaniu kryzysowym*. Ed E. Sobczak. Wydawnictwo Wydział Administracji i Nauk Społecznych Politechniki Warszawskiej, Warszawa.

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

SECURING AGENTS USING SECURE DOCKING MODULE

***Emil Gatial*[1], *Zoltán Balogh*[1], *Daniel M. Hein*[2], *Ladislav Hluchý*[1] *Martin Pirker*[2], *Roland Toegl*[2]**

[1] Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
[2] Institute for Applied Information Processing and Communications, Graz University of Technology, Graz, Austria

Key words: mobile agents, trusted computing, crisis management support.

Abstract

Modern communication and computing devices have the potential to increase the efficiency of disaster response. Mobile agents and seamless push-to-talk communication embody decentralised and flexible technologies to leverage this potential. While mobile agent platforms are facing greater variety of security risks compared to a classical client-server approach, trusted computing is capable of alleviating these problems. This document describes design and integration of a Secure Agent Infrastructure (SAI) with a Secure Docking Module (SDM) based on trusted computing principles for crisis management support. SDM provides a single chip security device that replaces the centralized trust decision and point with a suitable distributed solution. The main goal of SDM is protecting information. The protected information is only released to a requesting host device if the host is in a trusted state and adheres to a specific set of policies. SAI relies on the crypto-material protected by SDM thus the mobile agent can be unsealed only if the host machine is in the trusted state. The paper introduces the SDM and SAI technologies, describes motivation of SDM usage, provides summary of the key concepts behind the SDM and SAI. Further we provide analysis of requirements and security considerations as well as the integration points of the proposed architecture with other involved systems and the communication adapters between agents and other legacy systems. The last section concludes the article and presents our current achievements in integration and demonstration of the proposed technologies.

WYKORZYSTANIE SECURE DOCKING MODULE DO ZABEZPIECZANIA SYSTEMU AGENTOWEGO

***Emil Gatial*[1], *Zoltán Balogh*[1], *Daniel M. Hein*[2], *Ladislav Hluchý*[1] *Martin Pirker*[2], *Roland Toegl*[2]**

[1] Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
[2] Institute for Applied Information Processing and Communications, Graz University of Technology, Graz, Austria

Słowa kluczowe: mobilny system agentowy, trusted computing, zarządzanie kryzysowe.

Abstrakt

Współczesne rozwiązania teleinformatyczne mogą istotnie zwiększyć efektywność działań w sytuacjach kryzysowych. Systemy mobilnych agentów oraz „bezszwowa" komunikacja push-to-talk stanowią zdecentralizowane oraz elastyczne technologie wnoszące nową jakość do tej domeny. Rozwiązania oparte na mobilnych systemach agentowych są bardziej narażone na różnorodne zagrożenia w porównaniu z klasycznym rozwiązaniem klient–serwer; podatności te jednak mogą być redukowane dzięki zastosowaniu rozwiązań typu Trusted Computing. W artykule przedstawiono budowę oraz integrację Secure Agent Infrastructure (SAI) z Secure Docking Module (SDM) na podstawie zasad Trusted Computing. Rozwiązanie prezentowane w artykule jest przeznaczone do wsparcia zarządzania w sytuacjach kryzysowych. Głównym celem SDM jest ochrona informacji. Chroniona informacja jest udostępniania innym hostom tylko i wyłącznie, gdy znajdują się w stanie zaufanym oraz są w zgodności z określonym zestawem polityk.

W artykule opisano technologie SDM oraz SAI oraz uzasadniono stosowanie SDM. Przedstawiono także najważniejsze zagadnienia związane z SDM oraz SAI. Ponadto przeanalizowano wymagania oraz zagadnienia związane z bezpieczeństwem; wskazano także możliwości integracji zaproponowanej architektury z innymi systemami oraz urządzeniami komunikacyjnymi między agentami a tradycyjnymi systemami. W ostatniej części artykułu podsumowano jego treść oraz przedstawiono obecne osiągnięcia w dziedzinie integracji oraz demonstracji zaproponowanych technologii.

Introduction

Modern communication and computing devices have the potential to increase the efficiency of disaster response. Mobile agents and seamless push-to-talk communication embody decentralised and flexible technologies to leverage this potential. While mobile agent platforms are facing greater variety of security risks compared to a classical client-server approach, trusted computing (TC) is capable of alleviating these problems. Unfortunately, remote attestation, a core concept of TC, requires a powerful networked entity to perform trust decisions. The existence and availability of such a service in a disaster response scenario cannot be relied upon. One of the challenges of the communication infrastructures for distributed systems is to add new smart functions to existing services which would make the communication more effective and helpful for users. The aim is to provide smart functions via distributed IT systems which should provide a secure distributed paradigm to achieve confidentiality and access to resources. Such infrastructure should further provide a smart negotiating system for parameterization and independent handling of access requests to achieve rapid reaction. A good application of proposed system provides crisis management support that requires existing information from legacy systems of various organizations and from human operators in order to semi-automatically manage the crisis mitigation process or to enact decisions at various management levels. This information collection must be enacted in a secure manner while ensuring trust between both parties – information consumers and information providers. Many actors participate in a crisis situation, the competences and responsibilities of all parties are explicitly defined in a crisis mitigation plan. Information gathering is enacted either from legacy systems or from human end-users through mobile devices by guided dialog.

Several crisis response systems have been successfully built using multi-agent paradigm and other systems are being developed. Systems like DrillSim (Balasubramanian 2006), DEFACTO (Marecki 2005) and Mobile-FIRST (Honda 2009) were developed to simulate disaster situation using software agents enabling human actors to act more effectively. More realistic deployment of agent system was developed in the ALADDIN project (Jennings 2010) demonstrating the usefulness of decentralised and autonomous agent behaviour in the disaster management domain. Voyager (2011) communication platform delivers highly collaborative, dynamic, cross-platform applications and infrastructure for all business situations without the need of overwhelming modification of underlying corporate information systems. Specific use of mobile agents was presented in VEMPR system (Martin-Campillo 2009) dealing with reliable access to medical records of victims and in PA-UWNT research project (Kopena 2005) managing communication in mobile ad-hoc network project and Web-service based applications.

In this article we focus mainly on the concepts of security and trust used in Secure Agent Infrastructure (SAI) developed in the scope of SECRICOM integrated EU project (SECRICOM 2012). The goal of presented SAI is to enable easy collaboration and information sharing among actors in crisis situation, with an emphasis on security and trust of the information. In the following chapter, we present the architecture of SAI communication platform that deals with secure and trusted data collection during the crisis mitigation. We describe concepts of Secure Docking Module (SDM) and Trusted Computing approaches establishing trusted computing environment for SAI. Final part is devoted to description integration of SAI and SDM and to description of testing infrastructure. We conclude with achievements of SAI and SDM integration.

Architecture Design

We present a distributed architecture designed for the management of crisis situations where multiple actors are involved from various organizations with different competences and communicating over IP-based networks including wireless. We decided to design and implement such an architecture using agent paradigm. The distributed agent-based infrastructure is designed as a collection of software services with agent-like features (such as code

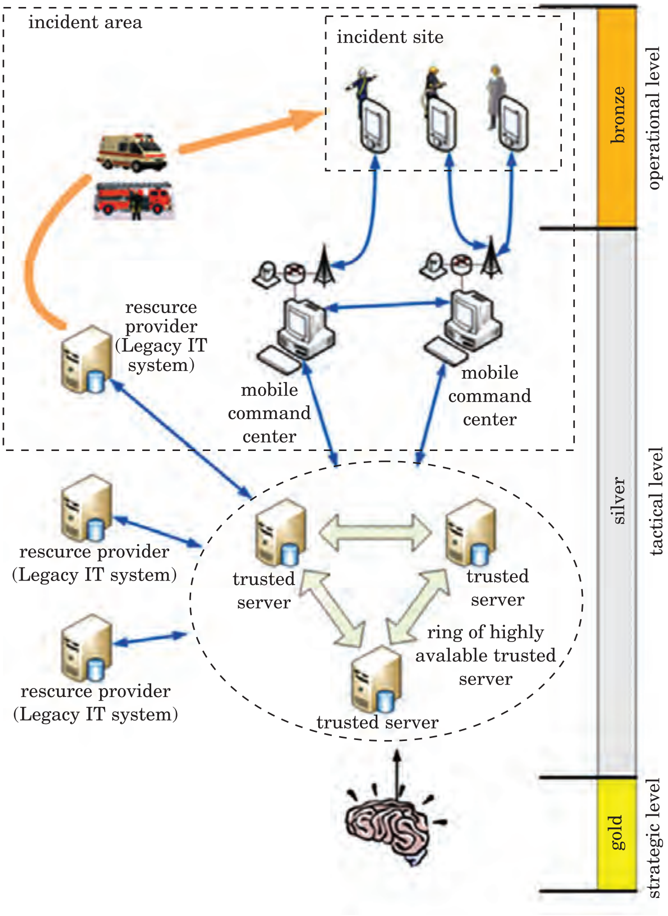


Fig. 1. An overview of Secure Agent Infrastructure applied in crisis management scenario

mobility) which would execute in a secure and trusted manner. Agent technology was selected due to the ability to fulfill such requirements through support of mobile and dynamically deployable executable code. Other advantages of agent-based systems are that they can help overcoming temporal or longer term communication network failures, save network bandwidth by being executed remotely and deliver only the execution results, provide means to execute code on remote host platforms in a trusted and secure manner or deploy code on host platforms on demand. The role of agents in the architecture is primarily coordinated collection of information. Information gathering is enacted either from legacy systems or from human end-users through mobile devices by guided dialog. With respect to requirements the overall agent infrastructure must be a secure, robust and failure resistant system. Because validity and authenticity of gathered information is a key factor for decision making in crisis management, trust must be set between agents and third party information systems. Also, agents must trust the host platform providers. The required level of trust for agents is based on a special hardware module – SDM providing TC functionality.

The home platform for agents is a network of Trusted Servers (TS) as it is depicted in the above figure (Fig 1). There are many different users involved in crisis management. Each type of user has a different level of responsibility, performs different tasks and requires different information (CRADDOC 2008). Gold Commanders who are in charge of producing strategy require information about the incident and about its effects on the wider area. They rarely need to make instant decisions, so have some time available to absorb information. Silver and Bronze Commanders are usually located closer to an incident site and need more detailed information about the incident and the resources available to them, as they have to turn the Gold-level strategy into a response, but are not as concerned with events outside the incident. They may have to make quick decisions as events unfold. Response Team Commanders and responders who are implementing a response have limited time in which to take in information and, as such, only need information relative to their immediate task. The coordination of responders; actions as well as providing live information to commanders in Silver and Gold level are the most important challenges in crisis management.

Concept of Docking Station Functionality

The SDM should allow agents to dock on a secure communication infrastructure by ensuring the state of the device it is supporting. The SAI is a distributed system and operates on confidential data. Therefore, the system must protect its integrity against data loss/theft and data modification. In a distributed system, data protection concerns are not limited to data transmission. As the data are processed in different physical computing platforms it must be established that all data processing entities adhere to the same security policy for the data. The data security policy adherence is enforced by ensuring the software configuration of a computing platform before it is connected to the SECRICOM infrastructure. To this end the SDM protects communication keys and credential information and only releases this information to the host platform if this platform is in an approved software configuration. The process of establishing the fact that a platform has an approved software configuration is called local attestation verification. Conceptually, the SDM protects a small set of key pairs for asymmetric cryptography, but in general is capable of protecting arbitrary data up to a specific size. The SDM's key protection facilities are a standard function, which could already be implemented with today's smart cards or hardware security modules. The SDM extends this standard function by only releasing these keys to a host device if and only if this host device is in a trusted state. This host device is called Trusted Docking Stations (TDS). The relationship between SDM and TDS is depicted in the figure below (Fig. 2).

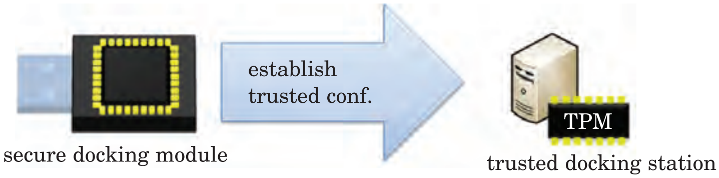


Fig. 2. Relationship between Secure Docking Module and Trusted Docking Station

A trusted platform software configuration is a specific software configuration. This software configuration is measured by a Trusted Platform Module (TPM). The combination of a SDM with a TDS is called a Secure Docking Station (SDS).

Trusted Computing

Generally, TC approaches were summarized in the work (PEARSON 2002). Trusted computing as specified by the Trusted Computing Group (TCG 2007) enables the authentication of a computing platform's software configuration. The software configuration is measured and mapped to a single value. The authenticity of this value is corroborated by signing it with a unique private key. This process is called attestation. Attestation allows a verifying entity to establish the software identity of a platform and correlate it with a configuration that enforces a set of required policies. If a platform's software configuration adheres to this set, we refer to this software configuration as trusted software configuration. For the attestation process to be valid, the software configuration measurements must be protected against tampering, the private signing key must be protected against misuse and compromise. Also, the private signing key must be bound to the measured platform. For these reasons, the core component of TC is a trusted module which fulfills these requirements. The components of the architecture can be broken down into different blocks, namely Secure Boot, Base System, Trust Management and Virtualization Partitions. The following figure (Fig. 3) illustrates these blocks.

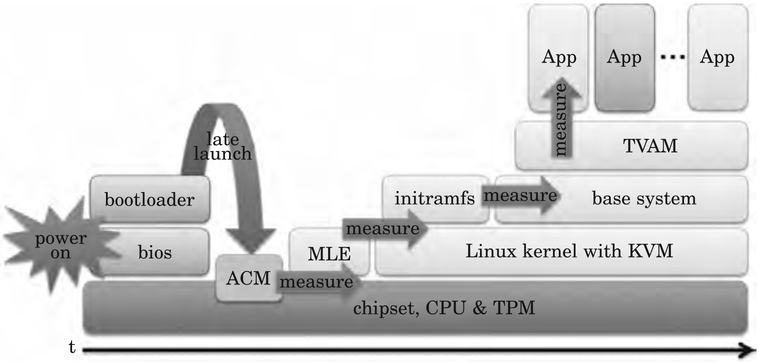


Fig. 3. Overview of the main components of the platform: Secure Boot, Base System, Trust Management and Virtualization Partitions. Trusted components are in green, untrusted are in red. The timeline indicates the different phases of platform boot

The Secure Boot block is responsible for initializing the system to a predefined configuration that requires close cooperation of hardware and software modules. We use Intel TXT as physical platform (Intel TXT 2011). The software side is accomplished by using a standard boot loader (GRUB) along with SINIT and tboot. Upon power-on, the platform performs a conventional boot, but does not start an operating system; instead, the MLE is prepared and a TXT late launch is performed. The precise, desired software configuration is specified by the administrator in the form of policies stored in the TPM. The LCP is evaluated by SINIT and specifies which MLE is allowed to be executed. tboot's policy is called Verified Launch Policy (VLP), and it contains known-good values for measurements of the Linux kernel and its temporary ram disk initramfs. A secure boot is performed into a hardware guaranteed state and the chain of trust is extended over the kernel and initramfs. If the measurements do not match the expected values provided by the VLP, tboot will shut the platform down. The startup code in the initramfs ensures an unbroken chain-of-trust; it measures the file system image of the full Base System into a PCR before it is mounted.

The Base System is a customized Linux operating system. The kernel is augmented with the Kernel-based Virtual Machine (KVM) hypervisor module. KVM requires common Commodity PC platform equipped with virtualization extensions. KVM can run multiple virtual machines, where each virtual machine has private virtualized hardware like a network card, hard disk, graphics adapter, etc. Those virtual devices are forwarded to QEMU (QEMU

2012), fast software which emulates a full hardware platform. To support deterministic PCR measurement, the Base System's file system must remain read-only. A temporary file system provides the needed read-write storage during platform operation. However, changes to the Base System do not survive a reboot of the platform. This ensures robustness of the base system image to malicious modifications. Management of the virtual partitions itself is done by a component called TVAM, the Trusted Virtual Application Manager. Virtualization Partitions may host any system normally running stand-alone. This can be an unmodified out-of-the-box Linux or Windows system, or a heavily customized system.

Securing Agents in Trusted Environment

The SAI actually provides the software components (Hector 2005) needed to run agents. Moreover, TDS uses SDM to setup a TC environment and thus enforces the policies required by the legacy systems. SDM releases the protected cryptographic material if and only if the TDS was booted into the trusted state; that means the platform is in the well known state. The DSAP service employs the SDM for storing the TDS private key, which is used to decrypt incoming agent's symmetric key to be run in a trusted environment.

The root of trust is established between the agents' home platform and host platform (HP) by audited agent code before its usage will take place. The audit process must ensure that the agent does only what its creator states it should do, and that it does not contain any malicious code, which may jeopardize the integrity of the HP. Establishing the trust between an agent and a HP is depicted in the next figure (Fig. 4).

Agent repository (AR) holds the set of certified agent Java classes or jar files. The code of agents may vary from executing simple DB query to complex management of HP resources. It is up to the agent designer to implement an agent's functionality, but with respect to the fact that the code must be audited and certified whether by the HP provider or by a trusted third-party authority. Based on the code certification the HP provider can trust the code running his or her HP. When Process Management Subsystem (PMS), which is specialized system coordinating data collection, decides to issue an agent it queries AR to obtain the classes implementing the agent. Here, PMS is able to verify the certificate of agent classes. Next, an instance of agent object is created by PMS where the agent attributes are set. The agent object and its classes are encrypted using an AES key secured by $TDS_1PubK_{E/D}$ public key (referred to as key encapsulation) (PSEC-KEM 2008) of HP. After the encrypted agent is moved on the HP, the DSAP service decrypts the AES key using $TDS_1PrK_{E/D}$

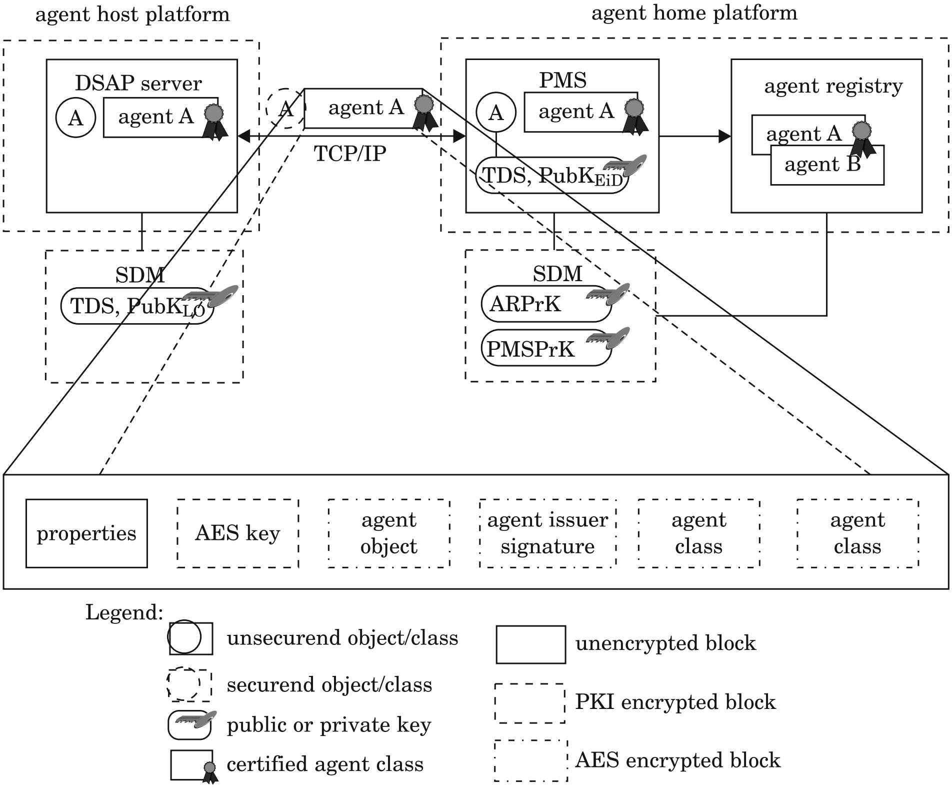


Fig. 4. The scheme of DSAP concept to establish secure and trusted communication of agents

private key of the HP (received from SDM) and uses this key to decrypt an agent. The HP usually provides access to some resources that a specific agent is able to process. Here, PMS is responsible for choosing the right type of agent and for setting him up to provide the required results. The results are encrypted using the same AES key and sent back to PMS.

Testing Infrastructure

The coordination of agents in SAI platform was tested in the scenario of free hospital beds reservation, while rescuing injured people. The infrastructure, (Fig. 5) comprises four fictive hospital information systems, where each system is attached to DSAP platform secured by SDM module (Linux OS). Next, the dedicated host platform running PTT client (Windows OS) is included in order to support end users to communicate with SAI via PTT

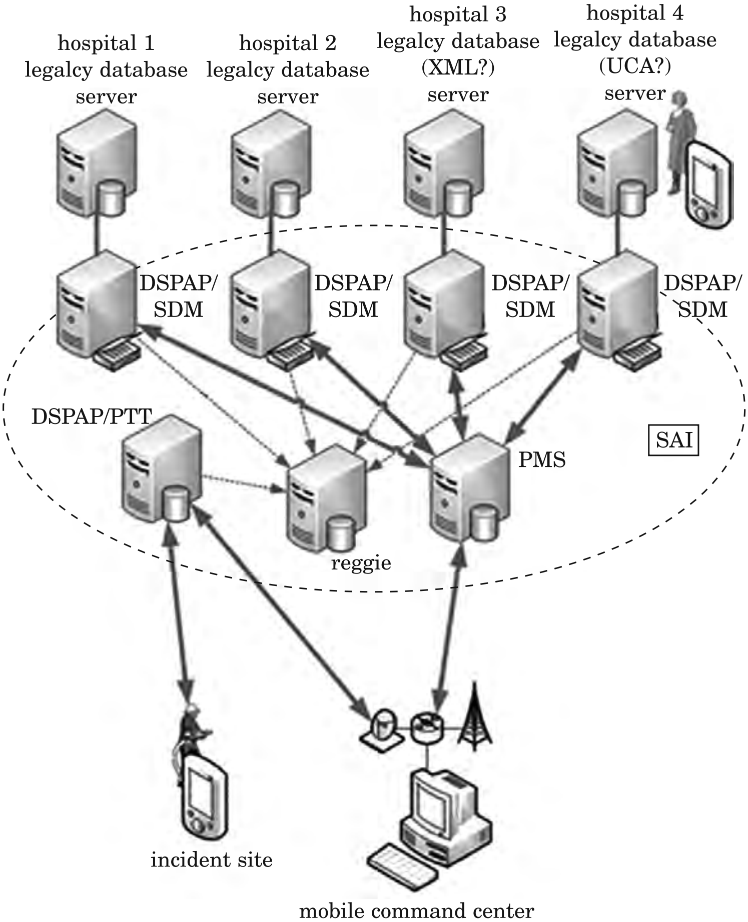


Fig. 5. SAI testbed infrastructure

enabled devices. Reggie component contains registry of DSAP services available in the network. Finally the specialized component called PMS coordinates agent communication and deployment.

The process of SAI enabled crisis management support starts when a first responder needs to find the closest hospital with a particular type of injury treatment. He can directly specify the type of process in the PTT enabled device or call the command centre. User communication agent (delegated by PMS) then collects information using guided dialog requesting the injury type, number of injured people and the position of incident (by location name or by GPS coordinates). By submitting the request the PMS sends information delivery agents to every nearby hospital system to query specific data and send them back to PMS. PMS then automatically reserves specific number of free hospital beds which are closest to incident location.

Conclusion

In this paper we described the integration of secure agents with a secure communication infrastructure for rapid information gathering in a crisis situation. Requirements for using secure agents arose from communication challenges in crisis management problem domain. The concept of SAI shows big potential in the applications using data from different legacy information sources or even different end-users using different communication channels. Moreover, the applications can benefit from the agent mobility and TC by processing data at hosting storage element or in its vicinity. The benefits of SDM as opposed to attestation based on sealing are twofold. First sealing is rather inflexible and does not allow easy migration. The SDM on the other hand can be plugged into any device with the appropriate interface. Furthermore, it is simpler to maintain a set of valid platform software configurations on the SDM, because it represents a single point of management. The second reason is that the SDM is a physical token. Its possession alone contributes to the authentication of the owner and it cannot be plugged into two devices at once. This restricts access to one device at the time.

Acknowledgment

This work is supported by project SeCriCom FP7-218123. This publication is the result of the project implementation: "Industry research in the area of effective work with large data in user oriented applications, ITMS code: 26240220029" and "RPKOM, ITMS code: 26240220064" supported by Operational Programme Research & Development funded by the ERDF.

Translated by Authors

Accepted for print 30.06.2012

References

Craddock R. 2008. *The UK Civilian Command and Control Hierarchy for Crisis Management, Responsibilities and Information Flow.* Thales Research and Technology (UK) Limited.

Balasubramanian V., Massaguer D., Mehrotra S., Venkatasubramanian N. 2006. *DrillSim: A Simulation Framework for Emergency Response Drills.* Proceeding ISI'06 Proceedings of the 4th IEEE international conference on Intelligence and Security Informatics, pp. 237–248.

Hector A., Narasimhan V.L. 2005. *A New Classification Scheme for Software Agents.* Proceedings of the Third International Conference on Information Technology and Applications (ICITA'05), IEEE Computer Society, ISBN:0-7695-2316-1, pp. 191–196.

HONDA J.M. 2009. *Application of Mobile Agent Systems to First Responder Training*. MSc. Thesis, University of California, http://www.cs.ucdavis.edu/research/tech-reports/2009/CSE-2009-13.pdf.

INTEL TXT. 2011. *Intel® Trusted Execution Technology (Intel® TXT)*. Software Development Guide, March, http://download.intel.com/technology/security/downloads/315168.pdf.

JENNINGS N.R. 2010. *ALADDIN End of Year Report*. Southampton, UK: University of Southampton, http://www.aladdinproject.org/wp-content/uploads/2011/02/finalreport.pdf.

KOPENA J., SULTANIK E., NAIK G., HOWLEY I., PEYSAKHOV M., CICIRELLO V.A., KAM M., REGLI W. 2005. *Service-Based Computing on Manets: Enabling Dynamic Interoperability of First Responders*. Journal IEEE Intelligent Systems Archive, 20(5).

MARECKI J., SCHURR N., TAMBE M. 2005. *Agent-based simulations for disaster rescue using the DEFACTO coordination system*. Wiley, pp. 2–19.

MARTIN-CAMPILLO A., MARTI R., ROBERTS S., GARCIA C.M. 2009. *Mobile Agents for Critical Medical Information Retrieving from the Emergency Scene*. In 7th Internacional Conference on Practical Applications of Agents and Multi-Agent Systems.

PEARSON S. 2002. *Trusted Computing Platforms: TCPA Technology in Context*. Published by Prentice Hall, ISBN-10: 0-13-009220-7.

PSEC-KEM. 2008. *PSEC-KEM Specification version 2.2*. NTT Information Sharing Platform Laboratories, NTT Corporation, April 14.

QEMU. 2012. *Quick EMUlator*. http://en.wikibooks.org/wiki/QEMU.

SECRICOM. 2012. *SECRICOM FP7 integrated project*. http://www.secricom.eu/.

TCG. 2007. *TCG Specification Architecture Overview*. Specification Revision 1.4 2nd August.

TPM. 2007. *Trusted Platform Module*. TCG TPM specification; Version 1.2; Revision 103, https://www.trustedcomputinggroup.org/specs/TPM/.

Voyager. 2011. *200Voyager Pervasive Platform*. http://recursionsw.com/Products/voyager.html#.

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

SECURE DOCKING STATION AND ITS PROTECTION AGAINST HARDWARE ATTACKS

Apostolos P. Fournaris[1,2], Jacques Fournier[3], Daniel Hein[4], Guillaume Reymond[3]

[1] Electrical and Computer Engineering Dpt, University of Patras, Rio Campus, Greec
[2] KNOSSOSnet Research Group, Informatics and Mass Media Dpt, Technical Educational Institute of Patras, Greece
[3] CEA-LETI Minatec, Gardanne, France
[4] Institute of Applied Information Processing and Communications, Graz University of Technology, Graz, Austria

Key words: Hardware, Security Module, Physical attacks, AES, RSA, security counter – measures.

Abstract

Security and Trust in communication systems where very sensitive information are exchanged is achieved and retained through hardware means. In the SECRICOM project where seamless, interoperable crisis management communication is required, we have developed a security and trust managements mechanism based on a smart card like hardware structure called Secure Docking Module (SDM). However, given the highly secure and hostile environment (emergency, crisis situation) where the SDM needs to function, this security module can be the subject of many attacks. While cryptanalytic attacks on the SDM security are impossible due to the employed strong cryptographic algorithms, attacks targeting the SDM implementation constitute a pragmatic threat that cannot be neglected. In this paper, we address possible hardware issues of the SDM chip and focus on the Hardware attack protection mechanisms especially on the SDM RSA and AES cryptographic accelerators. We present the research work that was done through the SECRICOM project on the above issues and analyze the basic concept behind the protected RSA-AES structures that complement the SDM architecture. Those hardware structures are fully compatible with the SDM protocols and offer strong protection against hardware power attacks and fault attacks while retaining high performance characteristics.

MODUŁ SECURE DOCKING STATION ORAZ JEGO OCHRONA PRZED ATAKAMI SPRZĘTOWYMI

Apostolos P. Fournaris[1,2], Jacques Fournier[3], Daniel Hein[4], Guillaume Reymond[3]

[1] Electrical and Computer Engineering Dpt, University of Patras, Rio Campus, Greece
[2] KNOSSOSnet Research Group, Informatics and Mass Media Dpt, Technical Educational Institute of Patras, Greece
[3] CEA-LETI Minatec, Gardanne, France
[4] Institute of Applied Information Processing and Communications, Graz University of Technology, Graz, Austria

Słowa kluczowe: mechanizmy bezpieczeństwa, Secure Docking Module (SDM), ataki sprzętowe.

Abstrakt

Bezpieczeństwo i zaufanie w systemach łączności, gdzie są przetwarzane informacje niejawne, jest zapewniane za pomocą rozwiązań sprzętowych. W projekcie SECRIOM, w którym jest wymagana interoperacyjna oraz „bezszwowa” łączność w zarządzaniu kryzysowym, wytworzono mechanizm zapewniania bezpieczeństwa oraz zaufania oparty na rozwiązaniu typu kart inteligentnych – Secure Docking Module (SDM). Biorąc jednak pod uwagę wysoki poziom zagrożenia środowiska łączności w sytuacjach kryzysowych, sam moduł SDM może być przedmiotem wielu ataków. Pomimo że ataki kryptoanalityczne na SDM są niemożliwe ze względu na zastosowane silne algorytmy kryptograficzne, zagrożenie wynikające z ataków na implementację SDM nie powinno być zaniedbywane. W artykule opisano możliwe problemy rozwiązań sprzętowych w chipie SDM oraz wyeksponowano mechanizmy zapobiegania atakom sprzętowym, szczególnie skierowanym na SDM RSA i akceleratory kryptograficzne AES. Zaprezentowano ponadto struktury RSA-AES, które uzupełniają architekturę SDM z punktu widzenia wzmocnienia ochrony. Te struktury sprzętowe są w pełni kompatybilne z protokołami w ramach SDM i oferują silną ochronę przed atakami fizycznymi, jednocześnie nie obniżają wysokich właściwości użytkowych.

Introduction

Trust in Information Systems constitutes a fundamental security issue in most sensitive data handling applications. However, achieving a high level of trust in the entities of such systems is not an easy task. There are some computer communication systems where the nature of the handled information is so sensitive that untrusted behaviors cannot be tolerated. In such systems, security and trust is ensured by hardware means.

SECRICOM European project is based on the efficient, seamless communication of civil emergency responders in situations of crisis (SECRICOM, 2008). The project's goal is to provide to emergency agencies a communication infrastructure that is fully interoperable regardless of what devices (mobile phone, smart phone, PC, Tablet, Push-to-Talk equipment e.t.c.) or communication system each agency uses (analog radio, GSM, 3G, TETRA, wifi, Internet e.t.c.).

In such a communication environment, strong security places a very important role. The communication channel where the all transactions are made should always remain secure and protected from eavesdropping and involved emergency responders must have trust to the communication they are engaged in and also trust that their device provides them with accurate, untampered data meaning that it has not been compromised. While there are several hardware means of achieving high security, very few solutions exist when it comes to offering trust. Trusted Computing Group's TPM chip is the most promising such solution however this Hardware Security Module (HSM) is not easily deployable in crisis situations where extreme conditions are at hand (portability of actors, various communication means, frequent disruptions of communication channels e.t.c.). TPM requires remote attestation procedures in order to guarantee a high trust level. In the crisis management case, that cannot and may not be provided, it would be much better if trust attestation is provided locally. The above reasons stemmed the need for a local security and trust attestation mechanism so within the SECRICOM project we designed a passive smart card like hardware token, complementing the TPM functionality and acting as a local trusted third party, capable of storing security keys, credentials and attesting the trust level of devices connected to it.

The SDM is described as an SD, MMC card or usb token that is physically attached to a Host machine and upon request from its host, releases the keys related to this host only if the host provides sufficient credential that it is in a trusted state. The keys for each host along with the host's id and public key are stored in the SDM secure memory. The SDM has a unique id number and a set of cryptographic keys (public, private key pair) that should not be transmitted in any way through the communication channel. Apart from the above, the SDM holds a series of valid configuration states (PCR values) for each Host authorized to communicate with it in order to be able to verify the trust state of such host.

In a way, the SDM plays the role of a local trusted third party. The process of verifying the host's trust level is called local attestation since it is similar to remote attestation but do not require a network communication channel since the SDM is attached to the host device (the attestation is performed locally). In an SDM enabled environment, the various programs of the network are controlled by trusted servers and are cryptographically secured using specific keys.

Based on the above concept, the SDM is capable of validating the local software integrity of a Host platform through trust measurements and providing sufficient proof that the measurements are authentic, fresh and untampered. The SDM as an add-on structure on the user's communication device is associated to a specific user through a password mechanism and therefore can bind a user along with the device to the crisis management communication system. As such, the fundamental security principle of user non-repudiation is retained, the user is bonded to the SDM and cannot deny its actions.

The SDM secret information that are handled by its cryptographic algorithms (RSA and AES) cannot be deduced using traditional cryptanalysis since the above algorithms are considered highly secure especially for high bit length keys (in the SDM case, 2048 bit RSA and 256 bit AES is used). However, there exist a series of attacks that do not target the cryptographic algorithm itself but the algorithm's implementation that can be successful in deducing secret data. Even if secret information cannot be learned, attackers may be able to disrupt the SDM hardware or deny service leading to other kinds of failures in the SECRICOM security system. Those Hardware attacks are powerful yet easy to mount and can be invasive, semi-invasive and non invasive. While invasive attacks require considerable expertise, chip depackaging and special equipment (laser cutter microscope, probes) in order to work, semi-invasive and non invasive attacks can be mounted by even inexperienced attackers following instructions or with cheap equipment. For the above reasons, in the SDM system design special care must be taken in order to include resistance against the above hardware attacks.

In this paper, we elaborate on the hardware structure of the SDM chip and focus on possible ways of attacking the SDM structure through hardware means especially by targeting the RSA and AES cryptographic accelerators. We present the research work that was done through the SECRICOM project on the above issues including Fault attack and side channel attack protection mechanisms capable of thwarting hardware attacks. The proposed hardware structures are fully compatible with the SDM protocols offering strong protection against power attacks and fault attacks while retaining high performance characteristics.

The remaining of the paper is organized as follows. In section 2, the SDM architecture is discussed briefly. Section 3 provides a general overview on Hardware attacks that can be mounted on the SDM and section 4 focuses explicitly on attacks on the SDM AES and RSA accelerators along with implemented countermeasures. Section 5 concludes the paper.

SDM Hardware structure

We envision the SDM as a synchronous System on Chip (SoC) device. The hardware structure of the SDM can be determined by the functions that it must fulfill. The SDM due to its potential connection with a TPM has a TPM-like structure and includes an RSA signature unit, a control processor unit, a non volatile memory unit for key storage, a true random number generator unit (TRNG), a SHA-1 hash function unit and a symmetric key encryption/decryption and key generation unit (AES algorithm).

The generic hardware structure of the SDM chip is presented in Figure 1. The system is structured around a data bus where all the data values are transferred for reading by or writing to a requesting unit of the SDM. There is also an address bus connected to the memory unit for a successful memory data reading and writing. An additional bus is also connected to all the units of the chip which is responsible for passing all the control signals to those units. Signals of this bus are in general managed by the processor.

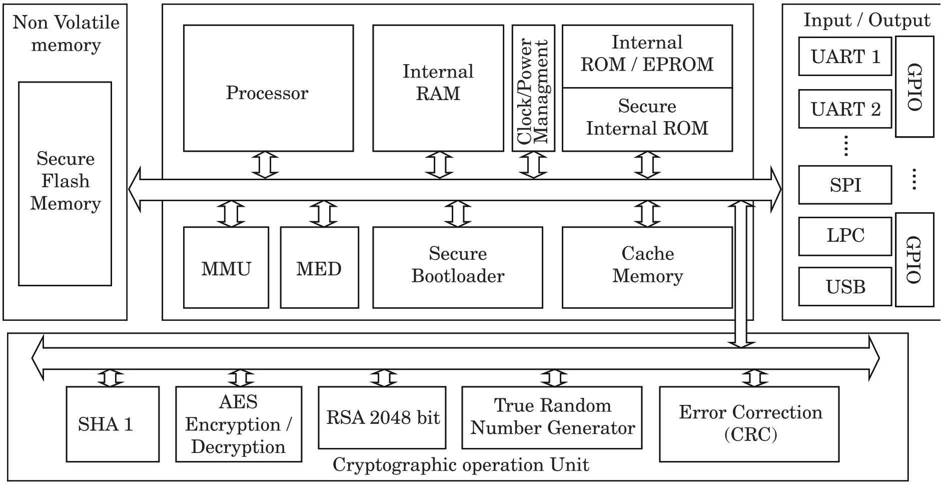


Fig. 1. The SDM hardware structure

The processor unit is responsible for controlling the whole SDM system and realizing the SDM functionality by enabling-controlling (chip select, CS, signal) SDM component units and performing operations that don't require the involvement of other units (i.e. comparisons or memory search). For this task, the processor has stored in its ROM memory a firmware program implementing the various SDM commands and a series of data required by those commands.

There are three memory modules included in the SDM hardware structure. The first module is a RAM unit that is employed for temporary value storage during a single SDM to Host protocol session. The second module is a ROM – EEPROM unit that is mainly used for storing and updating the SDM firmware realizing the SDM functionality. The third module is an NVRAM (flash memory) unit which constitutes the main storage area for all the sensitive information involved in the AAT should be SDM transactions. The three memory units are protected by special hardware structures that deter attackers from deciphering memory data.

The SHA-1 unit is implementing the SHA-1 hash function and the RSA encryption – decryption unit is responsible for performing the arithmetic operation of modular exponentiation ($m^e modN$) as defined in RSA public key scheme. The SHA-1 unit has a data input/output and a control signal indicating the beginning of a hash function operation. The RSA unit has as inputs the modulus N value (part of the RSA public key), the message m to be encrypted-decrypted and the public or private key e along with a control signal indicating the beginning of a modular exponentiation encryption or decryption.

The AES encryption/decryption unit is responsible for the key generation, encryption and decryption of an established session's data that are transmitted to and from the SDM. It has control signals that indicate an encryption, decryption and key generation operation. The AES generated session key is only saved within the AES unit and is changed when is replaced by a newly generated session key.

The TRNG unit is connected to the data path through its data output and has a 9 bit control signal that determines the bit length of the generated random number. The NVRAM has a read/write control signal while is connected to the data and address bus in order to read the address and use it to write or read the data values in or out of it. The system has a clock generator unit for managing the SDM different clocks and a reset signal along with chip select (CS) signals to enable or disable each SDM unit. The system works in a synchronous way.

Note, that in order to ensure a high security level, the RSA keys used in the SDM have a bit length of 2048 bits. As a result, the data related to the RSA encryption and decryption will have similar bit length. However, there is no feasible processing system able to operate with buses of such bits. Therefore, the 2048 bit values are broken into several blocks (to match the bus bit length) and reconstructed inside the SDM units (in example the RSA encryption-decryption unit). The same problem exists with the SHA-1 unit that handles 160 bit values and is solved in a similar way.

Hardware Attacks on SDM

The SDM can be considered as an embedded system that operates in hostile environment from security perspective. It can be stolen and manipulated in order to give out the sensitive information that are stored or processed in it. While the SDM system is protected from cryptanalytic attacks by the use of strong security schemes (RSA and AES), it can still be compromised when an adversary applies a hardware attack, an attack on the implementation itself. Three types of such attacks can be discriminated.

1. Invasive attacks. These attacks aim at physically disrupting the correct operation of an embedded chip. They involve removing chip packaging, micro probing of the chip's activity (memory, registers, buses) and physically tampering – interfering with the chip functionality. Invasive attack techniques begins by chip depackaging, removal of on-chip protection layers (depassivation) and modifying with probing tools the executed code or change values in Registers or simply observing the behavior of static/dynamic RAM blocks after power off since such units tend to "remember" values long after no power is applied to

them. Invasive attacks are not easily mounted, require attackers with considerable on chip expertise and expensive, specialize equipment like laser cuter microscope, micro probes e.t.c. Such attacks are considered difficult to be mounted on the SDM chip.

2. Semi invasive Attacks. These attacks aim at observing the behavior of the embedded system chip after an attacker specialized triggering. Like invasive attacks, they require depackaging the chip in order to get access to its surface. However, the passivation layer of the chip remains intact, as semi-invasive methods do not require depassivation or creating contacts to the internal lines. The goal is to induce a fault in the computation flow of the chip during a cryptographic operation and observe the cryptographic result as the fault propagates. The attacker can deduce sensitive information from such result on an unprotected embedded chip. In these attacks, also called Fault attacks, faults are injected using power or clock glitches, extreme variations in temperature, UV radiation or even optical laser beam induction. Depending on the attacker's equipment and expertise, the fault attacks are moderately difficult to mount. Fault attacks constitute a real danger for the SDM security.

3. Non-invasive Attacks. Such attacks, also called side channel attacks, exploit an embedded system's hardware characteristics leakage (power dissipation, computation time, electromagnetic emission e.t.c) to extract information about the processed data and use them to deduce sensitive information (cryptographic keys, messages e.t.c). An attacker does not tamper with the chip in any way and needs only make appropriate observations to mount a successful attack. Side channel attacks can be mounted very easily, cheaply, using a PC, a digital oscilloscope and some probes. Therefore, they can be mounted to an embedded system device by even the most inexperienced attacker. This ease of use makes SCA very potent. Some of the most widely used side channel attacks are the following:

– Power Attacks. These attacks involve physical measurement of the power dissipation emitted from the chip during cryptographic operations. Simple power signal analysis can reveal what mathematical operation is performed in the chip (e.g. modular multiplication or squaring during an RSA operation) and since in most cases the operation is related to the secret/private keys this action can reveal the key itself. Even if the chip is protected against simple power analysis, it is not fully resistant against power attacks since differential power analysis can still lead to compromise. In differential power analysis, the attacker perform guesses about a secret/private key bit, collects the related to this hypothesis power signal and correlates it with the actual power signal. The strongest correlation between the hypothesis and the actual measurement is the correct guess. Taking enough power samples and correlations the secret/private key can be revealed.

– Electromagnetic attacks. These attacks use the electromagnetic radiation emitted from the embedded system chip for simple analysis or differential analysis in a similar way as power attacks. In electromagnetic attacks, specific chip areas can be targeted and no physical access to the chip is strictly required (they can be mounted from afar). On the other hand, the existence of physical noise, RF interference or measurement error limits the attack's effectiveness.

Countermeasures

Designing countermeasures for Hardware attacks is not an easy task. Each security and cryptographic algorithm has its own vulnerabilities when implemented to hardware and therefore a ubiquitous approach toward Hardware attack resistance is impossible. In general, two approaches for countermeasures are used in practice, algorithmic based countermeasures and circuit based countermeasures. Algorithmic countermeasures aim at modifying the cryptographic algorithm and associated computer algebra operations so that when implemented it will leak to an attacked as less information as possible. Such countermeasures are more focused to semi-invasive and non-invasive attacks. Circuit countermeasures are hardware structures added to a cryptographic algorithm's hardware architecture, implementation or packaging, capable of detecting or thwarting a hardware attack. Such countermeasures can be used to protect an embedded system against all kinds of hardware attacks.

Invasive attack resistance is achieved by designing special structures during chip packaging and assembling in order to provide tamper evidence, detection and resistance. This may include mesh sensors implemented in the metal layer after packaging consisting of serpentine patterns of ground and power lines that are shortcircuited if attempts on depackaging or depassivation are done thus destroying the chip. Also, the on chip silicon layers can be designed in such a way that visual chip surface analysis through microscope is very difficult. Adding multiple layers with metal layers in between is such a technique applied during chip fabrication. All the invasive attack countermeasures are circuit based countermeasures.

Semi-Invasive attack resistance can be achieved by using some of the countermeasures for thwarting chip depackaging attempts (used in invasive attack resistance), however, usually such countermeasures are not enough or are too expensive. So, semi-invasive attack countermeasures are focused on detecting fault injection during cryptographic algorithm execution. One approach toward this end is to modify the cryptographic algorithm so as to support infective computation. The basic concept of infective computation is that any computational errors introduced by a fault will propagate throughout the cryptographic computation, thus ensuring that the final result appears random and useless to the attacker in the end. Another approach that can be combined with infective computation is the design of specialized fault detection units in the cryptographic algorithm hardware architecture capable of detecting single or multiple faults. Such units involve elegant circuit design as well as modifications in the cryptography algorithmic flow to include specific conditions between intermediate values that the fault detection unit must detect after the computations are concluded but before the cryptographic result is released. When faults are detected then a random number or zero value is released thus denying an attacker any useful information about secret/private keys.

There is a wide variety of non-invasive countermeasures depending on what side channel attack they thwart. The basic goal of all countermeasures is implementing the cryptographic architecture in such a way that the implementation's characteristics like power consumption, timing or electromagnetic radiation, leaks as little as possible of the secret keys or data. This can be achieved either by scrambling the leakage signal in such a way that is unrelated to the secret information that it is computed in the cryptographic unit or by minimizing the leakage as a whole so that it is very difficult for an attacker use it for a side channel attack. The first approach is related to algorithmic countermeasures that aim at inserting randomization through the cryptographic algorithm computation flow by providing Boolean, or arithmetic (multiplicative or additive) masking of the secret information (multiplication or addition with a random number). These countermeasures, also known as blinding, is very useful against Differential attacks since they aim at decorrelation of the secret data with the leakage itself. The second approach is related to algorithmic and mostly circuit countermeasures. Through special circuitry, like double rail technique, power rebalancing or additional dummy operations (redundancy), we aim at normalizing the leaked signals so that they remain unchanged during cryptographic operations. In general, it should be mentioned however, that protection against side channel attacks is never expected to be absolute: a determined attacker with a vast amount of resources can eventually, given enough time and effort compromise an implementation. The goal from cryptographic engineering perspective is to realize in the cryptographic accelerator enough side channel attack countermeasures so that an attack on the system becomes too expensive in effort or cost to be interesting (MANGARD et al. 2007).

Designing Protection Measures against SDM Side channel Attacks and Fault Attacks for AES and RSA

AES Accelerator countermeasures

AES accelerator implementation of the SDM can be the target of several fault and side channel attacks. Side channel attacks of special interest are differential attacks like *Differential Power or EM Analysis* (KOCHER et al. 1999, GANDOLFI et al. 2001, QUISQUATER et al. 2001), *Correlation Power Analysis* or *Mutual Information Analysis* (BRIER et al. 2004, GIERLICHS et al. 2008). Very potent AES implementation fault attack is *Differential Fault Analysis* (DFA) (BIHAM, SHAMIR 1997, PIRET et al. 2003, GIRAUD 2005).

Specific countermeasures on AES resistance against side channel attacks consist in either adding noise to blur the measurements or reducing the information-rich signal. "Balancing" consists in rendering the Hamming Weight (HW) or Hamming Distance (HD) of sensitive internal data constant by using "dual-rail" (each bit is encoded onto two wires) with, say, a "Return-to-Zero" (RTZ) (TIRI et al. 2003, SOARES et al. 2008, AMBROSE et al. 2011, CHEN et al. 2010). The propagation of the encoded values between the different parts of the circuit can also be physically balanced by using ad hoc Place and Route (P&R) techniques (GUILLEY et al. 2005). Moreover, noise can be added by randomizing the order of the instructions, by adding dummy operations or by masking the internal computations (AKKAR et al. 2001, TOKUNAGA et al. 2009).

As already stated in the previous subsection, countermeasures against fault attacks consist either in detecting errors during the computation and then taking actions to protect data or in making the circuit less sensitive to fault injections. The detection of error is mainly based on information redundancy either in space or in time (BERTONI et al. 2002, KARRI et al. 2003, KARPOVSKY et al. 2004) and can be further enhanced by placing several sensors in order to detect abnormal modifications of the chip's environment (voltage, temperature, clock frequency, light, etc.). Once a fault has been detected, reactions may consist in stopping the communication with the outside and/or resetting parts of the running software or deleting the sensitive data etc.

However, existing countermeasures do not address both fault and side channel attacks. In the AES SDM accelerator we realize countermeasures that are meant to thwart both classes of attacks. To achieve this, we designed an architecture based on duplicated-complemented (also called "dual") data paths applied to the AES algorithm. The dual data paths balance the data HW and are also used to detect faults.

The AES protection against DFA consists of detecting faults and reacting on them. We use spatial duplication in order to achieve that, as suggested in

(MALKIN et al. 2005) and implement two instances of the algorithm working in parallel. By checking the consistency between the results of the two instances we can determine if a fault have been injected (the outcome of the two instances won't match). If a fault injection is detected then the AES architecture reacts by returning an error value instead of the correct result. This value is extracted by blurring the erroneous ciphertext with the scrambled value of the detected fault. More on the fault detection mechanism can be found in (JOYE et al. 2007).

The AES protection against side channel attacks consists on designing the two parallel instances of the algorithm in such a way that when a bit of each intermediate value is computed in one instance, the other instance computes the complementary value. This approach effectively scrambles the power signal and disassociates power dissipation with the AES processed information (DOULCIER-VERDIER et al. 2011).

RSA Accelerator countermeasures

In the RSA cryptographic scheme, three n-bit numbers are used, the public modulus N, the public key e and the private key d. Let $N = p \cdot q$, where p, q are secret prime numbers. Let also $e \cdot d = 1 \mathrm{mod}\ (p - 1)(q - 1)$. Assuming that m is the message to be encrypted (plaintext), the RSA encrypted outcome (ciphertext) is $c = m^e \mathrm{mod} N$ and decrypted outcome is $m = c^d \mathrm{mod} N$. CRT is usually used during RSA decryption since the bit length of the private key d is bound to be long. In CRT RSA, we compute $S_p = c^{dp} \mathrm{mod} p$ and $S_q = c^{dq} \mathrm{mod} q$, where $d^p = d \mathrm{mod}(p - 1)$ and $d^q = d \mathrm{mod}(q - 1)$. Then, the final result is computed following Gauss's combination algorithm, meaning

$$S = \mathrm{CRT}(S_p, S_q) = (S_p \cdot q \cdot q_i) + (S_q \cdot p \cdot p_i) \mathrm{mod} N \qquad (1)$$

Where:
$q_i = q^{-1} \mathrm{mod} p,\ p_i = p^{-1} \mathrm{mod} q.$

The main computation and side channel attack security bottleneck of the RSA structure is the modular exponentiation operation. This operation is realized as a iterative process of modular multiplications that by themselves have considerable computation cost. Also, the exponent in modular exponentiation, which is usually sensitive information (private key), determines the computation flow during the operation's execution sequence and needs to be fast enough in order to support the SDM functional and non-functional specifications (speed, hardware resources e.t.c.).

We propose optimizing and enhancing the modular multiplication operation, constituting the heart of modular exponentiation and mounting hardware and algorithmic SCA countermeasures on the modular exponentiation operation itself. To achieve the first goal, an optimized version of Montgomery modular multiplier, described in (Fournaris 2010), is adopted. This structure employs Carry-Save logic in all its inputs, outputs and intermediate results as well as constant value precomputation during the multiplication algorithmic flow, managing to reduce considerably both the time and space complexity of modular multiplication.

To achieve the second goal, we focus our counter measures on popular side channel attacks that can easily be mounted on RSA. Such attacks are Power attacks (especially simple power attacks) and fault attacks. RSA Side Channel Attack countermeasures can be generic, on circuit level, (more effective against power attacks) (Bhattacharya et al. 2008), (Tiri et al. 2006) or specialized, focused on specific cryptographic algorithms, on an algorithmic level. The second approach can be more effective since it utilizes techniques that better negate the RSA cryptoalgorithm's specialized SCA weaknesses. Our design adopts the second approach since it makes the proposed architecture Hardware agnostic, meaning that it can guarantee Side Channel Attack resistance regardless of the Hardware implementation technology.

In general, the target of RSA Fault and Power attacks is the modular exponentiation unit. The use of CRT, increases RSA vulnerability to such attacks, so strong countermeasures are needed for modular exponentiation protection. Simple Power attack (SPA) resistance is achieved by making the arithmetic operations during the exponentiation algorithm execution undiscriminated from an external observer (Joye, Yen 2003). RSA Fault attack countermeasures are based on techniques of detecting single fault injection and blocking further processing thus prohibiting the release of secret information (Aumuller et al. 2002).

Assuming that a fault is introduced during the first exponentiation (with modulus p), then the faulty output would be S'_p and the CRT reconstruction in (1) would be $\overline{S} = \mathrm{CRT}(S_p,S_q) = (S'_p \cdot q \cdot q_i) + S_q \cdot p \cdot p_i)\mathrm{mod}N$.

Knowing a legitimate CRT-RSA outcome S and a faulty one $\overline{S}$, one can find the secret prime q by calculating $q = gcd((S - \overline{S}), N)$. The deliberate insertion of a fault in the computation flow of one of the exponentiations constitute a fault attack, originally proposed by Boneh et al. (1997) and enhanced by Lenstra (1996) where no legitimate outcome is also needed, as can be observed by $q = gcd((\overline{S}^e = m)\mathrm{mod}N, N)$.

To thwart the above attack, we employ Girauld's technique of detecting faults (Giraud 2006). This approach uses Montgomery power ladder as an iterative modular exponentiation algorithm thus apart from fault detection it offers resistance against SPA. Two Montgomery power ladder calculations are performed, in parallel in each modular exponentiation round with different initial inputs each. The initial input of the second Montgomery power ladder is the first round's result of the first Montgomery power ladder. At the end of each round, two values are calculated, S and S_1, that have a known mathematical connection between them ($S_{\mathrm{o}} =^{o} m \cdot S_1 \mathrm{mod}(p \cdot q)$). This connection exists due to appropriate initialization at the beginning of the algorithmic flow. Fault detection is performed in the end of a modular exponentiation after CRT reconstruction by checking if the connection between S_{o} and S_1 is true. If this test fails then a fault attack is detected and the cryptographic processes are canceled. Thorough analysis revealed that the above technique is not completely fault attack resistant as observed by KIM and QUISQUATER (2007), since a carefully injected fault after the final Montgomery power ladder round and before the fault detection operation theoretically can damage the whole protection mechanism without being detected. So, to thwart this attack, the final Montgomery power ladder round result must be masked by adding a random number *a* that is to be removed after CRT reconstruction and fault detection. In that case, the correct (unmasked) RSA result is not revealed nor stored during the whole RSA computations only after fault detection test is

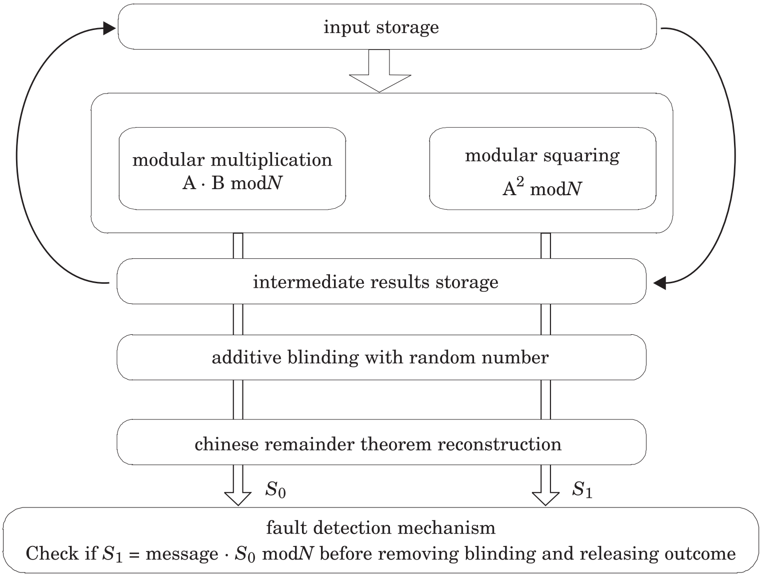

input storage
modular multiplication
A · B modN
modular squaring
A² modN
intermediate results storage
additive blinding with random number
chinese remainder theorem reconstruction
S0
S1
fault detection mechanism
Check if S1 = message · S0 modN before removing blinding and releasing outcome

Fig. 2. The SDM RSA accelerator concept and Hardware attack protection mechanism

passed. More on the SDM RSA accelerator can be found in (FOURNARIS 2010, FOURNARIS, HEIN 2011, FOURNARIS, KOUFOPAVLOU 2011). The concept behind the above approach is presented in Figure 2.

Conclusions

In this paper, the SECRICOM project hardware security and trust module (SDM) structure was analyzed and possible hardware attacks on its structure where mapped. The cryptographic accelerator structures (RSA and AES) of the SDM were identified as potential targets of such attacks and specific fault and side channel attacks on these structures were presented. Specific countermeasures both for side channel and fault attacks were presented for SDM AES and RSA implementations based on spatial duplication (for AES) and modified Montgomery Ladder technique (for RSA). The presented methodologies manage to protect the SDM structure from the most popular side channel and fault attacks and since they are based on open architectures (not proprietary), they can be expanded in the future to include resistance against attacks yet to appear.

Acknowledgment

This work was funded by the SECRICOM project (EC FP7-SEC-2007 grant 218123).

Translated by AUTHORS

Accepted for print 30.06.2012

References

AKKAR M.-L., GIRAUD C. 2001. *An Implementation of DES and AES, Secure against Some Attacks.* In: *Proceedings of CHES'01*. Edited by Çetin Koç, D. Naccache, C. Paar. LNCS, 2162: 309–318, Springer-Verlag, Paris, France.

AMBROSE J., RAGEL R., PARAMESWARAN S., IGNJATOVIC A. 2011. *Multiprocessor information concealment architecture to prevent power analysis-based side channel attacks.* Computers Digital Techniques, IET, 5(1): 1–15.

AUMULLER C., BIER P., FISCHER W., HOFREITER P., SEIFERT J.-P. 2002. *Fault Attacks on RSA with CRT: Concrete Results and Practical Countermeasures*, In: *Revised Papers from the 4th* International Workshop on Cryptographic Hardware and Embedded Systems (CHES '02), Eds. B.S. Kaliski Jr., C.K. Koç, Ch. Paar. Springer-Verlag, London, UK, pp. 260–275.

BERTONI G., BREVEGLIERI L., KOREN I., MAISTRI P., PIURI V. 2002. *A parity code based fault detection for an implementation of the advanced encryption standard.* In: *Proceedings of DFT'02*. IEEE Computer Society, Washington, DC, USA, pp. 51–59.

BHATTACHARYA K., RANGANATHAN N. 2008. *A linear programming formulation for security-aware gate sizing*. In: *GLSVLSI '08*. Proceedings of the 18th ACM Great Lakes symposium on VLSI. 1em plus 0.5em minus 0.4em New York, NY, USA: ACM, pp. 273–278.

Biham E., Shamir A. 1997. *Differential Fault Analysis of Secret Key Cryptosystems.* In: Proceedings of CRYPTO '97, LNCS, 1294: 513–525.

Boneh D., Demillo R.A., Lipton R.J. 1997. *On the Importance of Checking Cryptographic Protocols for Faults (Extended Abstract).* In Proc. EUROCRYPT, pp.37–51.

Brier E., Clavier C., Olivier F. 2004. *Correlation Power Analysis with a leakage model.* In: Proceedings of CHES 2004. Edited by M. Joye and J.-J. Quisquater, Lecture Notes in Computer Science, 3156: 16–29, Springer-Verlag.

Chen Z., Sinha A., Schaumont P. 2010. *Implementing virtual secure circuit using a custom-instruction approach.* In: Proceedings of CASES '10, ACM, New York, NY, USA, pp. 57–66.

Doulcier-Verdier M., Dutertre J-M., Fournier J., Rigaud J-B., Robisson B., Tria A. 2011. *A Side-Channel and Fault Attack Resistant AES circuit working on duplicated complemented values.* In: *Solid State Circuits Conference – Digest of technical papers, 2011 (ISSCC 2011).* Page 15.6, IEEE International.

Fournaris A.P. 2010. *Fault and Simple Power Attack Resistant RSA using Montgomery Modular Multiplication*. Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS 2010) IEEE, pp. 1875–1878).

Fournaris A.P., Hein D.M. 2011. *Trust Management Through Hardware Means: Design Concerns and Optimizations*. In: Eds., N. Voros, A. Mukherjee, N. Sklavos, K. Masselos, M. Huebner, Symposium VLSI 2010 Annual, vol. 105, pp. 31–45. Springer Netherlands.

Fournaris A.P., Koufopavlou O. 2011. *Efficient CRT RSA with SCA countermeasures.* In: Proceedings of 14th Euromicro DSD '11. Oulu, Finland, pp. 593–599.

Gandolfi K., Mourtel C., Olivier F. 2001. *Electromagnetic Analyis: Concrete Results.* In: *Proceedings of CHES'01*, Eds. Ç. Koç, D. Naccache, C. Paar, *LNCS*, 2162: 251–261, Springer-Verlag, Paris, France.

Gierlichs B., Batina L, Tuyls P., Preneel B. 2008. *Mutual Information Analysis – A Generic Side-Channel Distinguisher.* In: *Proceedings of CHES'08.* Eds. E. Oswald, P. Rohatgi, Lecture Notes in Computer Science, 5154: 426–442, Springer-Verlag, Washington DC,US.

Giraud C. 2006. *An RSA implementation resistant to fault attacks and to simple power analysis.* IEEE Transactions on Computers, 55(9): 1116–1120.

Giraud C. 2005. *DFA on AES.* In: *Advanced Encryption Standard – AES.* Eds. H. Dobbertin, V. Rijmen, A. Sowa, Lecture Notes in Computer Science, 3373: 27–41, Springer Berlin / Heidelberg.

Guilley S., Hoogvors T.P., Mathieu Y., Pacalet R. 2005. *The backend duplication method.* In: Proceedings of CHES'05, pp. 383–397.

Joye M., Yen S.-M. 2003. *The Montgomery powering ladder.* In: CHES '02: Revised Papers from the 4th International Workshop on Cryptographic Hardware and Embedded Systems. 1em plus 0.5em minus 0.4em London, UK: Springer-Verlag, pp. 291–302.

Joye M., Manet P., Rigaud J.-B. 2007. *Strengthening Hardware AES Implementations against Fault Attack.* IET Information Security, 1: 106–110.

Karpovsky M. G., Kulikowski K. J., Taubin A. 2004. *Robust protection against fault injection attacks on smart cards implementing the Advanced Encryption Standard.* In: Proceedings of DSN 2004, pp. 93–101, IEEE Computer Society.

Karri R., Kuznetsov G., Goessel M. 2003. *Parity-Based Concurrent Error Detection of Substitution-Permutation Network Block Ciphers.* In: Proceedings of CHES'03, LNCS, 2779: 113–124, Springer--Verlag, Cologne, Germany.

Kim C. H., Quisquater J.-J. 2007. *Fault attacks for CRT based RSA: New attacks, new results, and new countermeasures.* In: *WISTP*, Eds. D. Sauveron, C. Markantonakis, A. Bilas, J.-J. Quisquater. Lecture Notes in Computer Science, 4462. pp. 215–228, 1em plus 0.5em minus 0.4em Springer.

Kocher P., Jaffe J., Jun B. 1999. *Differential Power* Analysis. Advances in Cryptology Proceedings of Crypto 1999, pp. 388–397. Springer-Verlag.

Lenstra A. K. 1996. *Memo on RSA signature generation in the presence of faults*.

Malkin T. G., Standaert F.-X., Yung M. 2005./ *A comparative cost/security analysis of fault attack countermeasures.* In: Proceedings of FDTC'05, pp. 109–123, Edinburgh, UK.

Mangard S., Oswald E., Popp T. 2007. *Power Analysis Attacks: Revealing the Secrets of Smart Cards.* Springer.

Piret G., Quisquater J.-J. 2003. *A Differential Fault Attack Technique against SPN Structures, with Application to the AES and KHAZAD.* In Proceedings of CHES'03, 2 LNCS, 779: pp. 77–88, Springer-Verlag.

Quisquater J.-J., Samyde D. 2001. *Electromagnetic Analysis (EMA): Measures and coutermeasures for smart cards.* In: e-smart 2001, LNCS, 2140: 200–210.

SECRICOM. 2008. *Seamless communication for crisis management.* http://www.secricom.eu/menu-objectives.

Soares R., Calazans N., Lomné V., Maurine P., Torres L., Robert M. 2008. *Evaluating the robustness of secure triple track logic through prototyping.* In: Proceedings of SBCCI'08, pp. 193–198, ACM, New York, NY, USA.

Tiri K., Verbauwhede I. 2003. *Securing Encryption Algorithms against DPA at the Logic Level: Next Generation Smart Card Technology.* In: Proceedings of CHES'03, LNCS, 2779: 125–136, Springer-Verlag, Cologne, Germany, 2003.

Tiri K., Verbauwhede I. 2006. *A digital design flow for secure integrated circuits*, IEEE Trans. on CAD of Integrated Circuits and Systems, 25(7): 1197–1208.

Tokunaga C., Blaauw D. 2009. *Secure AES engine with a local switched-capacitor current equalizer.* In: *Digest of Technical Papers of ISSCC 2009.* IEEE International, pp. 64–65, San Francisco, USA.

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

SECURITY CHARACTERISATION OF A HARDENED AES CRYPTOSYSTEM USING A LASER

***Cyril Roscian*[1], *Florian Praden*[1,2], *Jean-Max Dutertre*[1], *Jacques Fournier*[2], *Assia Tria*[1,2]**

[1] ENSM.SE – École Nationale Supérieure des Mines de St-Étienne
[2] CEA-LETI Minatec, Gardanne, France

Key words: Advanced Encryption Standard (AES), fault attacks, laser, security characterization, DFA.

Abstrakt

The AES is a standard encryption algorithm used in numerous cryptographic systems like smart cards, TPMs as well as in protocols like WPA2 or OpenSSL. Measuring the robustness of AES implementations against physical attacks is of utmost import-ance in order to guarantee the security of the system into which the AES is used. In this article, we describe how a hardware AES, embedding countermeasures against physical attacks, has been characterized using a laser. With the latter, we tried to implement a class of physical attacks called fault attacks which, when successful, allows an attacker to retrieve the secret key used by the AES module. Our experiments have allowed us to validate the efficiency of some of the countermeasures implemented in this AES implementation and have given us hints on how to further improve them.

OKREŚLANIE POZIOMU BEZPIECZEŃSTWA SPRZĘTOWEGO MODUŁU AES Z WYKORZYSTANIEM LASERA

***Cyril Roscian*[1], *Florian Praden*[1,2], *Jean-Max Dutertre*[1], *Jacques Fournier*[2], *Assia Tria*[1,2]**

[1] ENSM.SE – École Nationale Supérieure des Mines de St-Étienne
[2] CEA-LETI Minatec, Gardanne, France

Słowa kluczowe: Advanced Encryption Standard (AES), wstrzykiwanie błędów, laser, badanie bezpieczeństwa, DFA.

Abstrakt

AES to standardowy algorytm szyfrowania stosowany w wielu systemach kryptograficznych, np. kartach elektronicznych, TPM, oraz takich protokołach, jak WPA2 czy OpenSSL. Pomiar odporności implementacji algorytmu AES na ataki fizyczne jest najważniejszy do zapewnienia bezpieczeństwa systemowi opartemu na AES. W artykule opisano, jak sprzętowa implementacja AES z wbudowanymi zabezpieczeniami przeciwko fizycznym atakom była badana z wykorzystaniem lasera. Następnie podjęto próby zaimplementowania ataków fizycznych polegających na wstrzykiwaniu błędów; ataki te – zakończone sukcesem – pozwalają atakującemu na przechwycenie tajnego klucza wykorzystywanego w module AES. Przeprowadzone eksperymenty pozwoliły na sprawdzenie efektywności zabezpieczeń zaimplementowanych w module sprzętowym AES oraz wskazały możliwości dalszego podniesienia poziomu bezpieczeństwa.

Introduction

Security, through the authenticity, confidentiality and integrity of communication systems, has become an essential component of all electronic systems. The vulnerability to attacks of the devices implementing the cryptographic algorithms (such as smart cards) has become a critical issue. Some malicious means or "physical attacks" could be used to retrieve sensitive information such as encryption keys and therefore lower the security of the whole protected transmission chain of information. There are three types of physical attacks: *invasive* attacks, which cover all the techniques based on the modification (Anderson 1998) and probing (Handschuh 1999, Schmidt 2009, Gammel 2010) of integrated circuits (IC) by an invasive method (Kömmerling 1999, Koeune 2005); *observation or passive* attacks which exploit the fact that some physical characteristics such as power consumption (Kocher 1999, Lu 2010), electromagnetic radiation or the duration of computation depend on the chip's internal calculations (Koeune 2005); *perturbation or fault* attacks, which are based on changing the environmental conditions of the chip to infer information about the internal state of the IC. The latter ones are the most complex to implement as various and complex parameters must be taken into account such as the timing (i.e. the synchronization between the injection time and the calculation), the localization and the power level.

In this paper we mainly discuss about security characterisations based on fault attacks. One of the goals of such attacks could be to reveal the secret keys of a cryptographic device based on techniques like Differential Fault Analysis (DFA) (Biham 1997, Boneh 1997). Several methods can be used to induce faults into cryptographic ICs: use of a laser (Skorobogatov 2005, Bar-El 2006), voltage pulses (Blömer 2003), clock glitches (Agoyan 2010) or electromagnetic (EM) disturbances (Schmidt 2007). The use of lasers is one of the most effective techniques. Lasers allow a good reproducibility, an accurate control on the timing of the injection (the instant of firing and the duration of the pulse) and a precise focalization (the ability to restrain its effect to a limited number of gates). Consequently, lasers generate precise local effects into the IC thus leaving the rest of the chip "undisturbed". Several theoretical attacks against cryptographic algorithms are based on such models of fault injection methods (Piret 2003, Giraud 2005, 2010, Moradi 2010). Although protections against these kinds of attacks exist (YEN 2006), advanced methods combine semi-invasive attacks and power or EM analysis (MORADI 2011).

In this article, we describe the security assessment of an AES hardware chip done to validate the efficiency of the embedded countermeasures that could be incorporated into the AES module of the SECRICOM'S secure docking module (SDM). The SDM contains authentication keys and access rights associated with each user of the SECRICOM secure communication network. To access such keys, a secure channel, based on AES encryption, has to be established between the SDM and its host device (Trusted Docking Module). Hence, guaranteeing the resistance of the AES used against physical attacks helps in hardening the security chain of SECRICOM's communication infrastructure.

The outline of this paper is as follows. First, we recall the physical phenomena triggered when a laser is used to inject a fault into an IC. Then we describe the AES implementation used as target of our laser-based characterization. Third, we provide a description of the laser-based tests made on the targeted AES. With attacks like those described in (PIRET 2003, GIRAUD 2005) in mind, we tested the countermeasure in a "black box" approach (i.e. without using the knowledge we had of the implemented countermeasures). We also tested the AES in a "white box" approach where we tried to take advantage of the knowledge we had of the implemented countermeasures in order to circumvent them. Finally we provide a discussion about the results that we obtained and the conclusions that could be drawn from them.

Physical phenomena induced by lasers into an IC

When analysing the effect of a laser beam onto an IC, two phenomena have to be considered: the photoelectric effect appearing and the fact that some parts of the IC are more sensitive to the laser than others. When a laser beam strikes the Silicon and that the photon;s energy is higher than the Silicon's band gap, electron-hole pairs are created. In general, these pairs recombine and there is no effect on the IC. However under specific conditions, some undesired effects can appear. We shall focus on one of those effects called the Single Event Effect (SEE).

Single Event Effect

A Single Event Effect can appear when the electron-hole pairs created by the laser beam are drifted in opposite directions by the electrical field in the PN junction instead of immediately recombining. The consequence of that is the creation of a transient current as illustrated in Figure 1.

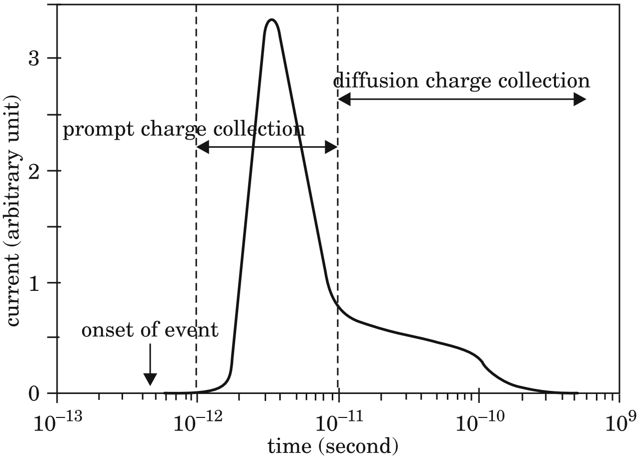


Fig. 1. Transient current resulting from charge collection after laser shoot

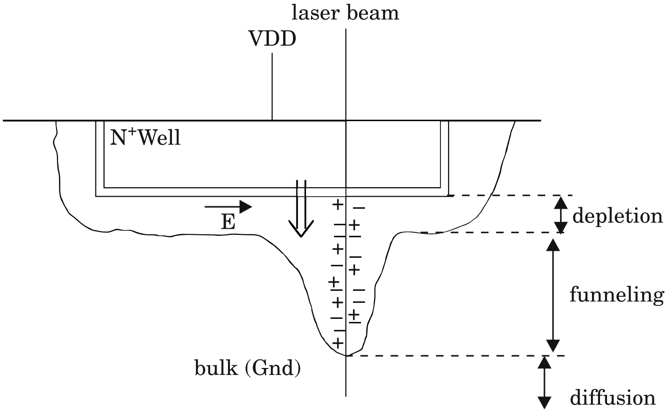


Fig. 2. Laser beam effect into the MOS structure

Along the laser beam shown in Figure 2, after the creation of the electron-hole pairs, two phenomena lead to the creation of the transient current. The first phenomenon stretches the depletion region (hence the extension of the electric field) along the laser beam and within a few picoseconds, charges are collected giving a current peak. In a second time, the rest of the charges are collected in a longer diffusing scheme. Figure 2 shows the transient current associated with the two phenomena of collection as given in (WANG 2008).

Sensitive zones

In CMOS technology, some parts are more sensitive than others to SEEs. To create an SEE, and then a fault, a strong electric field is needed. The reverse biased PN-junctions of the chip provide this required electric field. The position of these junctions can change, depending on the value of the data manipulated.

A good example to illustrate this data dependency is the CMOS inverter. The first case is a high state on the inverter's input: the NMOS transistor is in the "ON" state, its drain is grounded, the source and the bulk too and there is no reverse biased PN-junction. The PMOS transistor is in "OFF" state, then its source and the N well are in high potential, but its drain is grounded giving rise to a reverse biased junction. Hence, the drain of the PMOS transistor becomes sensitive to a laser shoot. In the same way, with a low state on the inverter's input, the drain of the NMOS transistor becomes sensitive to a laser shoot.

Figure 3(a) shows an inverter with a high state on its input and the "sensitive zone" is coloured in red. The second inverter (b) on the figure represents the other case of localisation of the "sensitive zone".

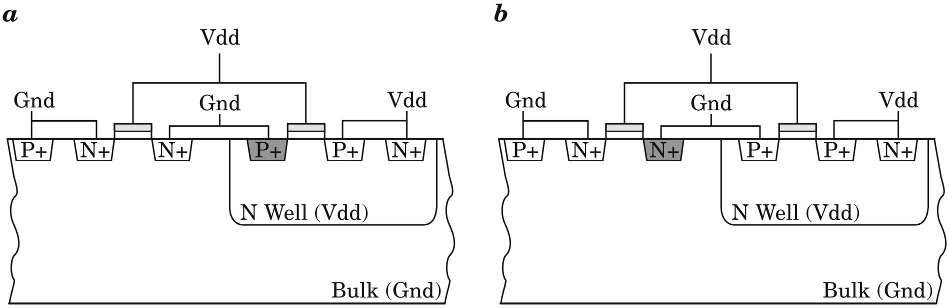


Fig. 3. CMOS inverter with a high state: *a* – or low state, *b* – on its input

From SEE to faults

We explained in the previous sections how a laser beam can create single events into CMOS structures and which parts of it are more sensitive to a laser. Even if an SEE is created by a laser, it is possible that the SEE has no effect on the chip's computations. An SEE can be transformed into a fault in two different ways. The first one is to generate an SEE directly into a register. In this case, the register's state is changed and this change is stored and propagated. The second way consists in creating, into the chip's logic, an SEE which propagates through the logic up to the next register. Depending on the timing, if the SEE reaches the register's input on a clock's rising edge, an "faulty" value will be latched. Thus, a fault is injected into the chip's computations. Figure 4 illustrates the propagation of an SEE into the logic and the difficulty of transforming it into a fault. In the first case, the SEE generated in the logic is not captured by the D flip flop of the register cell and has no significant effect on the data processed. In the second case, with the adequate timing, the SEE is captured by the D flip flop. Thus, the value of the register is changed: the SEE has been turned into a fault.

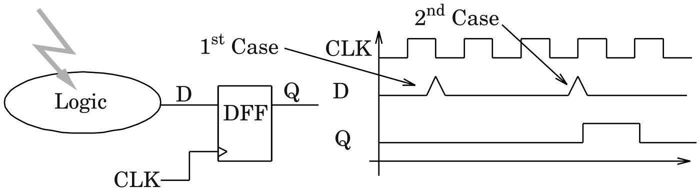


Fig. 4. Propagation of SEE through Logic

The hardened AES test chip

In this section, we describe the chip used as a device under test (DuT) for our laser-based characterizations. The DuT is a hardware module implementing the Advanced Encryption Standard (AES) algorithm used, for example, for encrypting the secure channel between a SECRICOM host device and the SECRICOM's SDM.

The AES algorithm

The AES algorithm is a symmetric key cryptography standard established by the NIST (NIST 2001). This algorithm is a substitution and permutation network based on four transformations (SubBytes, ShiftRows, MixColumns, AddRoundKey) used iteratively in rounds (Figure 5). In this paper, we focus on the 128-bit key version. This version processes data blocks of 128 bits, considered as matrices of 4x4 bytes called States, in ten rounds. The round keys (K1 to K10) used during every round are calculated by a key expansion routine (not detailed in this paper). We refer as M1 to M10 the States at the end of each round.

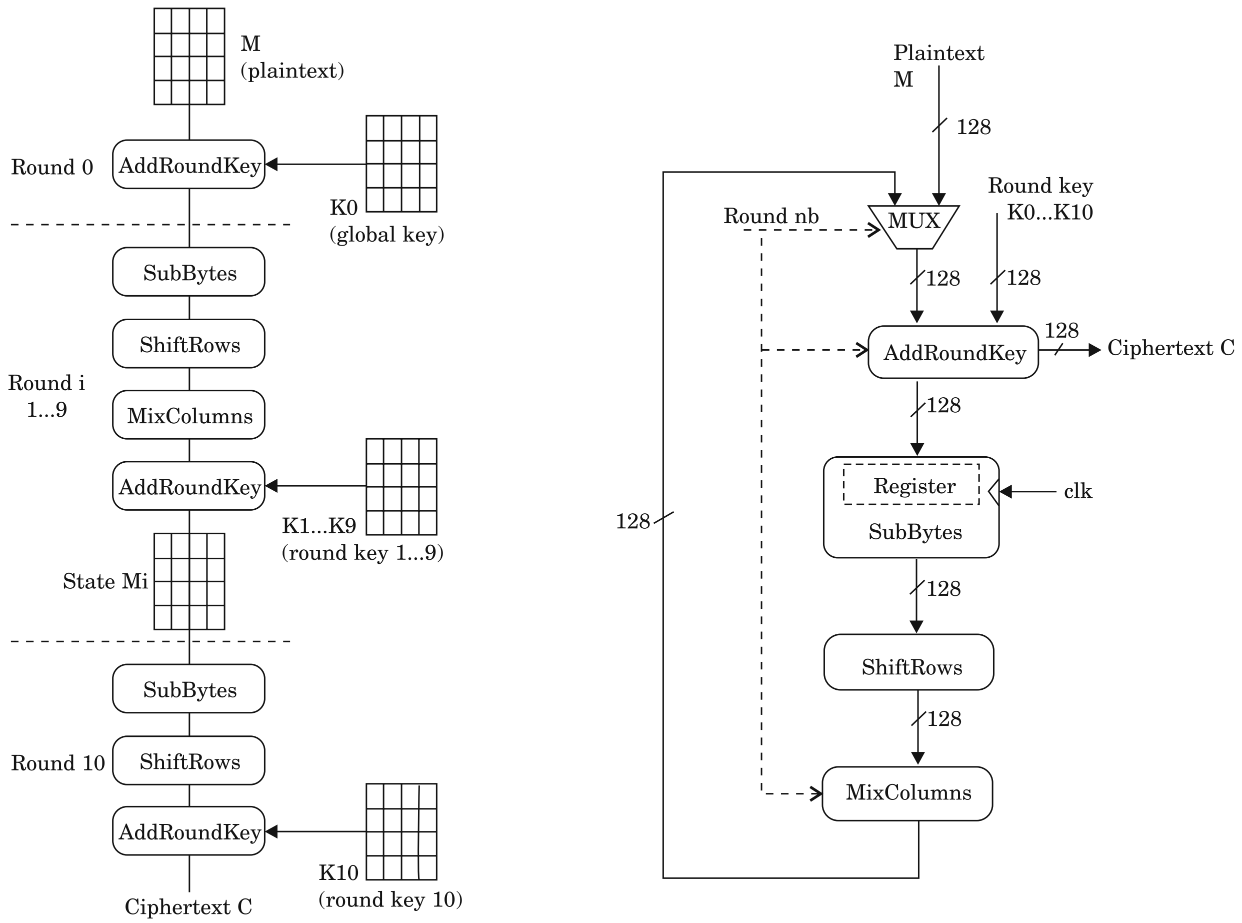


Fig. 5. The AES algorithm and its hardware implementation

The secure ASIC AES

In our study, we use the secure AES test chip described in (AGOYAN 2011) implemented in HCMOS9 gp 130 nm STM technology. The size of the die is 1336,m x 1411.8 μm and its working frequency is 25 MHz. A picture of the chip is shown in Figure 6.

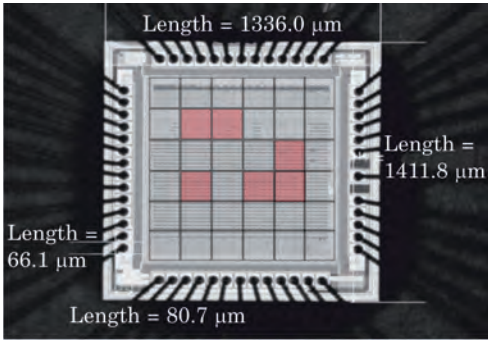


Fig. 6. Sensitive areas of the ASIC

The countermeasure against faults attacks implemented in this chip consists first in detecting errors and then in reacting in case of detection. The error detection is done by using spatial duplication: the AES is executed twice in parallel and at each round, the results of the two instances (the original path and the duplicated one) are compared. If an error is detected, the reaction consists in blurring the erroneous cipher text with a scrambled value of the detected error. The error detection mechanism is described in the following section.

Figure 7 illustrates the architecture of the implemented AES chip. We can see the two AES rounds executed in parallel with the error detection system.

The error detection mechanism

As mentioned in the previous section, when an error is detected (*XNOR* operation between the States from the two paths), the error detection mechanism scrambles the error value and then blurs the cipher text with the scrambled error. An error matrix is used whereby the error is spread across the rows and the columns as shown in the Figure 8. After that, the error matrix is XORed with the SubByte's results of the two data paths.

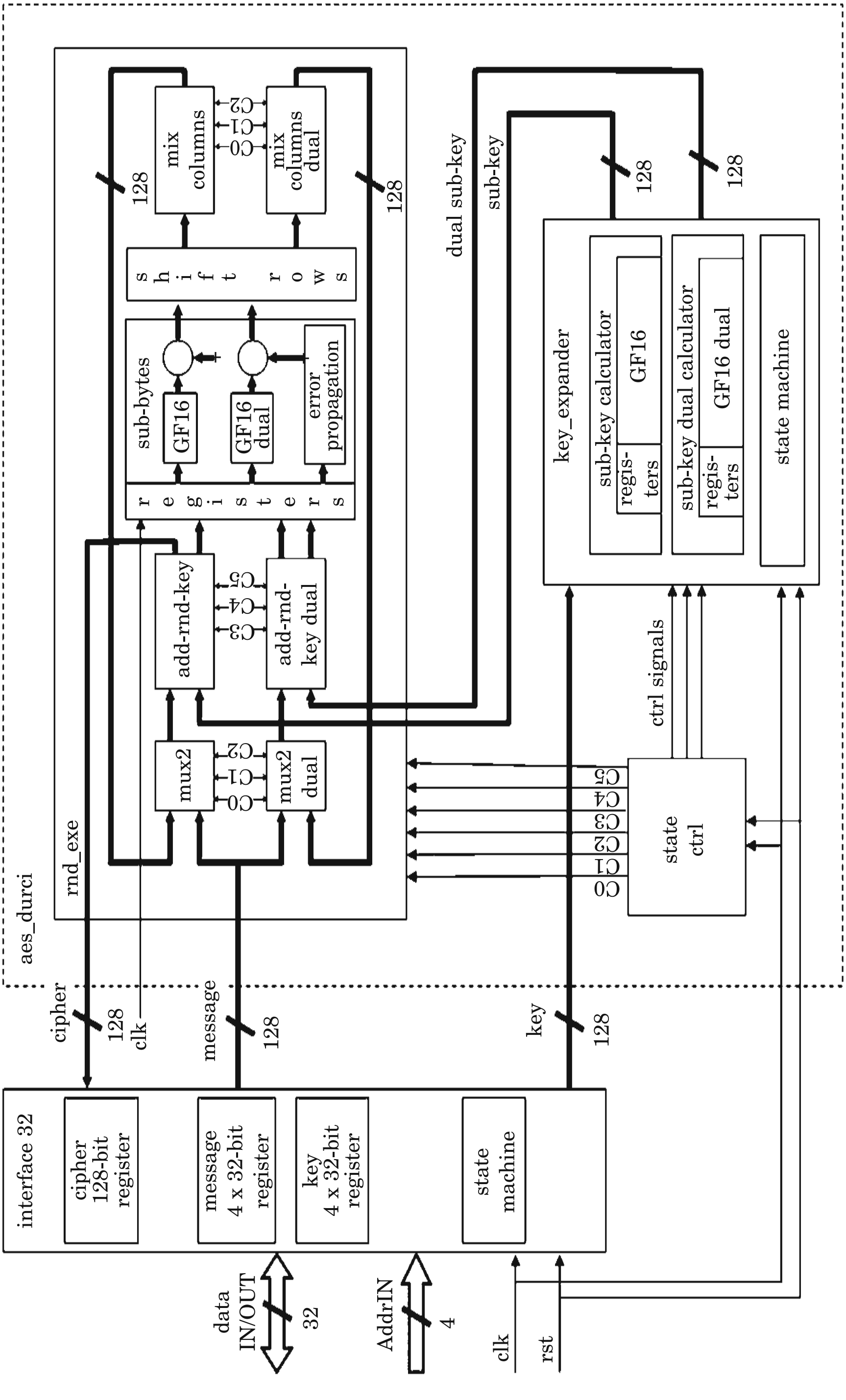


Fig. 7. Overview of the ASIC AES architecture

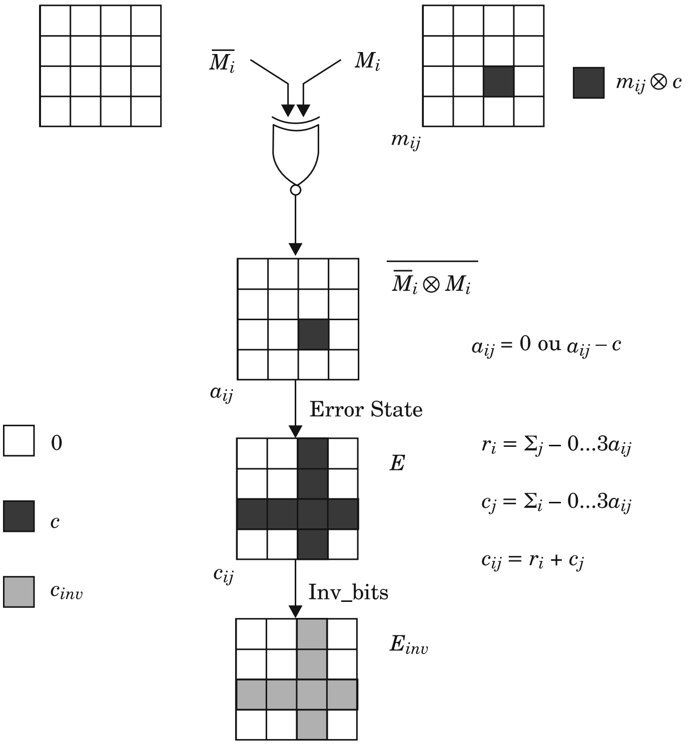

$\overline{M_i}$
M_i
m_{ij}
$m_{ij} \otimes c$
$\overline{\overline{M_i} \otimes M_i}$
$a_{ij} = 0$ ou $a_{ij} - c$
a_{ij}
Error State
0
c
c_{inv}
E
$r_i = \Sigma_j - 0...3a_{ij}$
$c_j = \Sigma_i - 0...3a_{ij}$
$c_{ij} = r_i + c_j$
c_{ij}
Inv_bits
E_{inv}

Fig. 8. The Error Detection and Spreading Mechanism

The Cross-ShiftRows operation

In addition to the above error detection and spreading mechanism, the ShiftRows operation is crossed between the two data paths. Half of the bits of each byte come from the other path and vice versa. This "crossing" operation is an additional security barrier to further scramble the error. If a fault is injected onto one of the paths, the fault is detected and propagated onto the two paths in parallel. With the Cross-ShiftRows, half of the information is lost due to the transfer to the other path. Figure 9 illustrates the injection of a fault on the last round of the AES and its propagation throughout the two data paths.

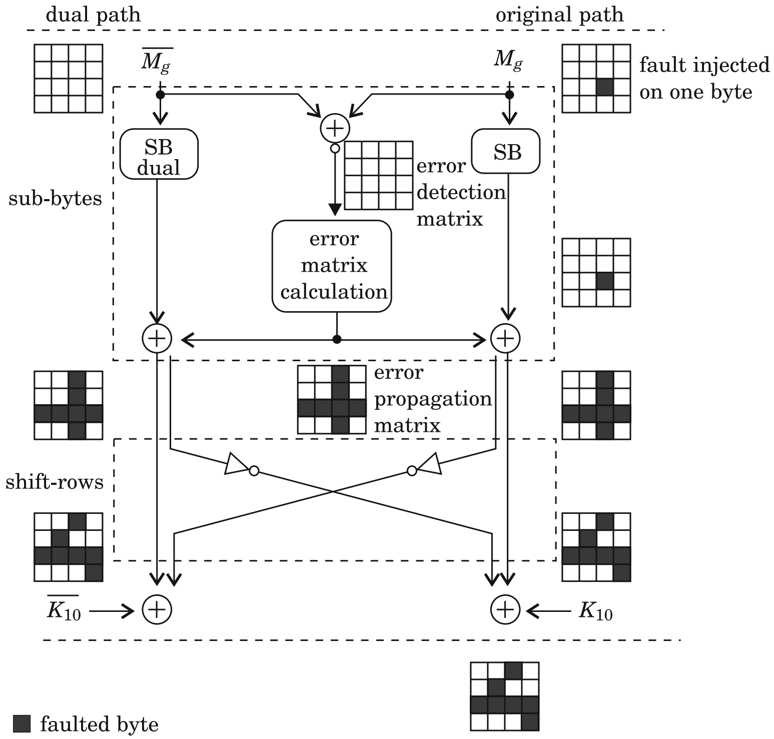


Fig. 9. Propagation of a fault on the last AES round (with the countermeasures)

Fault injection using laser

The Laser test bench

To perform the different tests on the AES chip, we used the laser facility of the MicroPackS$^{\mathrm{TM}}$ platform (Micropacks 2012). The laser used is a YAG (Yttrium Aluminium Garnet) laser with three different sources: green, infrared and ultra-violet. We used the green source (with a wavelength of approximately 532 nm) with a 20× Mitutoyo lens. We obtained a spot size between 1 μm and 150 μm. With the largest spot size, we have approximately 15pJ of energy per laser shoot.

The AES is interfaced with a control PC. When an encryption is launched, a trigger signal is sent to an FPGA synchronization board, which sends a shoot signal to the laser after a delay defined by the control PC. This delay allows triggering the laser at different times during the encryption calculation.

We put ourselves into two configurations when doing those tests. In the first configuration we adopted a "black box" approach where we ignored any of the implementation information we had on the DuT and tried to perform fault injections on the data path of the AES, in the “classical one”, with the objective of collecting the erroneous cipher texts and doing differential cryptanalysis like in Giraud’s or Piret’s methods. In the second configuration, we used our knowledge of the implemented countermeasures, in a so-called “white box” approach, to try to circumvent the security mechanism by trying to fault the detection mechanism itself.

Fault injection on the data path

The easiest way to generate errors and trigger the detection mechanism is to inject faults into the register of the SubByte module. Despite the propagation of the error into the two data paths and the loss of half of the information due to the Cross-ShiftRows, we can always try to use the faulty cipher texts in Giraud’s DFA (GIRAUD 2005). The *sine-qua-none* condition for this attack is to generate mono-bit faults (i.e. errors on only one bit of the State matrix). The error matrix can be found with a simple XOR operation between the correct cipher text and the faulty one. Figure 9 illustrates this kind of fault injection.

Fault injection on the detection mechanism

Another way of generating errors is to use the error detection mechanism itself. The DFA described in (PIRET 2003) needs a fault injection before the last MixColumns operation (Round 9). With our countermeasures and a fault injection on the data path, the faulty cipher texts cannot be exploited for this attack.

When looking closer at the Cross-ShiftRows operation, it appears that if the same fault is injected on the two data paths, the effect of the Cross-ShiftRows is “neutralized”. Due to the dispersion of the lay-out of the two paths across the chip’s surface, it’s very hard to inject the same fault into the two paths with a laser which has a local effect. The solution is to inject the fault directly into the error matrix. By doing so, we could propagate the same error into the two paths and the Cross-ShiftRows will be “neutralized”. In the last round, as the errors are the same on the two data paths, no detection appears and we have a faulty cipher text that could be used for DFA. Figure 10 depicts the propagation into the two paths of an error injected directly into the error matrix at the 9th round.

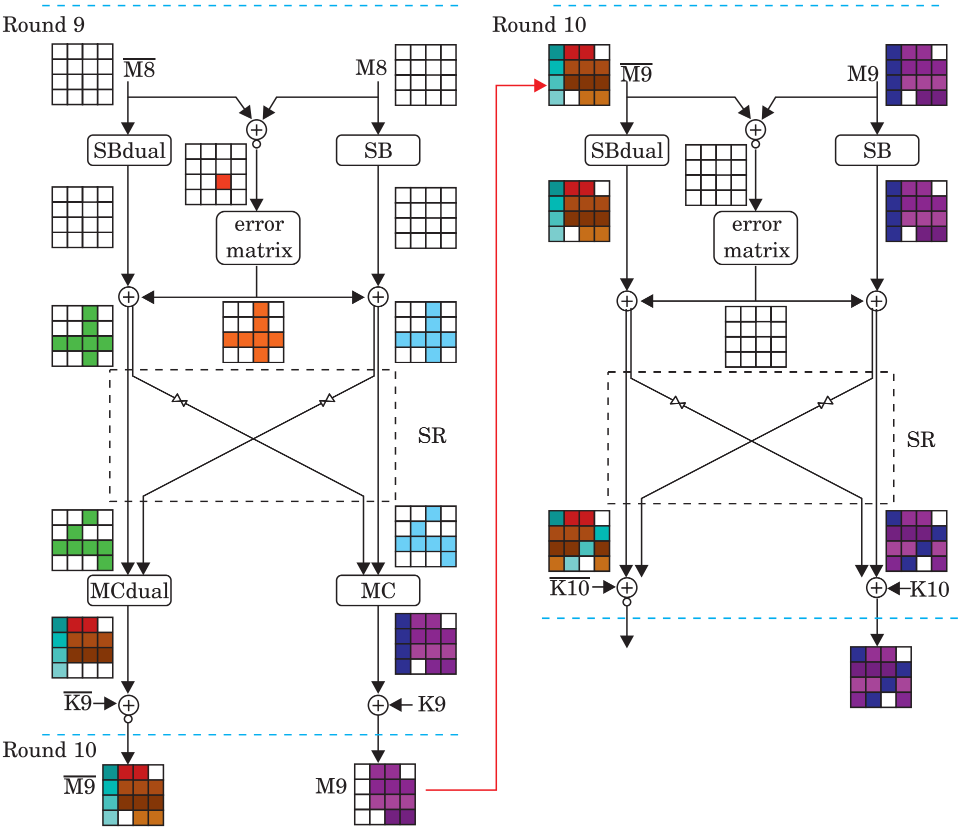


Fig. 10. Injection of an error into the error matrix and its propagation

Laser tests on the AES chip

Localization of the SubByte's registers

Since the registers of the SubByte are dispersed across the ASIC, the first step of the characterization work was to physically localize those registers. To do so, we made a cartography using the laser. The surface of the chip was partitioned into 36 zones of size 150 µm x 150 µm each as shown in Figure 6. For each zone, 50 different encryptions were performed among which we looked for specific faulty cipher texts: as shown in Figure 9, if a fault is injected on one byte at the beginning of the last round, in the end, we obtain the same error in six bytes of the resulting cipher matrix, due to the propagation mechanism, and one different error at the position of the injected fault. In Figure 6, the coloured regions highlight those where such specific faulty cipher texts were obtained and which appeared to be more sensitive to this type of fault on the last round of encryption.

Results

Once the SubByte's registers have been localized, we started injecting faults on the data path. In the "black box" approach, we tried to perform DFA as described in (GIRAUD 2005) using only the faulty and correct cipher texts but in vain: the detection and error spreading mechanism proved to be efficient against such attacks.

However, when we used the knowledge of the implementation of the countermeasure (i.e. in a "white box" approach), especially the structure of the Cross-ShiftRows, we managed to recover a few bytes of the secret key of the AES. The complete knowledge of the Cross-ShiftRows is necessary because half of the information is lost in this operation and we need, for the attack, to keep all the information.

We also tried to inject faults directly into the error matrix in order to try another type of DFA (PIRET 2003) where in theory two faulty cipher texts are needed to recover 4 bytes of the secret key. To find all the 16 bytes of the key, we need to inject an error in one of the bytes of each column of the error matrix. Despite our efforts, we couldn't inject any fault into the error matrix with our laser test bench. One of the reasons for this is that the error matrix is not implemented using registers but with logic gates. Thus it is very hard to synchronize the laser shoot with the ASIC's encryption precisely enough to target separately each column of the error matrix.

Discussion and Conclusion

In this paper, we have described how countermeasures implemented in a hardware implementation of the AES have been tested using a laser as fault injection means. We have seen that, in a "black box" approach, classical DFA techniques are inefficient against such countermeasures. However our characterization work has also shown that in a "white box" scenario, some bytes of the secret keys could be recovered. This has led us to the conclusion that the error propagation should have been truly random and independent from the generated errors (requiring the implementation of a True Random Number Generator in the chip). We also investigated another attack path by trying to inject a fault in the error matrix itself but this has been unsuccessful illustrating the limits of current equipment with respect to current technologies. Such characterization works have provided valuable design rules for implementing secure encryption AES modules like those used in the SDM of the SECRICOM project.

Acknowledgements

This work was funded by the SECRICOM project (EC FP7-SEC-2007 grant 218123). The research work of Cyril Roscian was partly funded by the "Conseil Regional Provence-Alpes Côtes d'Azur".

Translated by AUTHORS

Accepted for print 30.06.2012

References

AGOYAN M., DUTERTRE J-M., NACCACHE D, ROBISSON B., TRIA A. 2010. *When Clocks Fail: On Critical Paths and Clock Faults*. SPRINGER VERLAG ed. Smart Card Research and Advanced Application.

AGOYAN M., BOUSQUET S., DUTERTRE J-Max., FOURNIER J., RIGAUD J-B., ROBISSON B., TRIA A. 2011. *Design and characterisation of an AES chip embedding countermeasures*. International Journal of Intelligent Engineering Informatics, 1, 3–4: 328–347.

AMIEL F., CLAVIER C., TUNSTALL M. *Collision fault analysis of DPA resistant algorithms.* In the proceedings of Fault Diagnosis ąand Tolerance in Cryptography 2006 – FDTC 2006.

ANDERSON R.J., KUHN M.G. 1998. *Low Cost Attacks on Tamper Resistant Devices.* In the Proceedings of the 5th International Workshop on Security Protocols.

BAR-EL H., CHOUKRI, H., NACCACHE D, TUNSTALL M., WHELAN C. 2004. *The Sorcerer's Apprentice Guide to Fault Attacks*. E-Print: 100.

BIHAM E., SHAMIR A. 1997. *Differential Fault Analysis of Secret Key Cryptosystems.* In the proceedings of the 17th Annual International Cryptology Conference on Advances in Cryptology.

BLÖMER J., SEIFERT J. 2003. *Fault Based Cryptanalysis of the Advanced Encryption Standard (AES).* In the proceedings of Financial Cryptography.

BONEH D., DEMILLO R.A., LIPTON R.J. 1997. *On the Importance of Checking Cryptographic Protocols for Faults*. Advances in Cryptology – EUROCRYPT '97, International Conference on the Theory and Application of Cryptographic Techniques, Konstanz, Germany, May 11–15.

GAMMEL B.M., MANGARD S. 2010. *On the Duality of Probing and Fault Attacks.* J. Electron. Test., 26(4): 483-493 ISSN 0923-8174. DOI 10.1007/s10836-010-5160-0.

GIRAUD C. 2005. *DFA on AES.* In the proceedings of the 4th international conference on Advanced Encryption Standard. Bonn, Germany.

GIRAUD C., THILLARD A. 2010. *Piret and Quisquater's DFA on AES Revisited*. E-print: 440.

HANDSCHUH H., PAILLIER P., STERN J. 1999. *Probing Attacks on Tamper-Resistant Devices.* In the Proceedings of the First International Workshop on Cryptographic Hardware and Embedded Systems.

KIM C.H., QUISQUATER J-J. 2008. *New Differential Fault Analysis on AES Key Schedule: Two Faults Are Enough.* In the proceedings of the 8th IFIP WG 8.8/11.2 international conference on Smart Card ąResearch and Advanced Applications. London, UK.

KOCHER P.C., JAFFE J., JUN B. 1999. *Differential Power Analysis.* In the proceedings of CRYPTO.

KOEUNE F., STANDAERT F. *A Tutorial on Physical Security and Side-Channel Attacks.* In Foundations of Security Analysis and Design III: FOSAD 2004/2005, Nov 2006, 3655, 78–108

KÖMMERLING O., KUHN M.G. 1999. *Design principles for tamper-resistant smartcard processors.* In the Proceedings of the USENIX Workshop on Smartcard Technology on USENIX Workshop on Smartcard Technology. Chicago, Illinois.

Lu J., Pan J., Den Hartog J. 2010. *Principles on the security of AES against first and second-order differential power analysis.* In the Proceedings of the 8th international conference on Applied cryptography and network security. Beijing, China.

Micropacks. http://www.arcsis.org, last accessed 19th of April 2012.

Moradi A., Mischke O., Paar C., Li Y., Ohta K., Sakiyama K. 2011. *On the power of fault sensitivity analysis and collision side-channel attacks in a ącombined setting.* In the proceedings of the 13th international conference on Cryptographic hardware ąand embedded systems. Nara, Japan.

Moradi A., Shalmani M.T.M., Salmasizadeh M. 2006. *A generalized method of differential fault attack against AES cryptosystem.* In the Proceedings of the 8th international conference on Cryptographic Hardware and Embedded Systems. Yokohama, Japan.

Mukhopadhyay D. 2009. *An Improved Fault Based Attack of the Advanced Encryption Standard.* In the Proceedings of the 2nd International Conference on Cryptology in Africa: Progress in Cryptology. Gammarth, Tunisia.

NIST, National Institute of Standards and Technology. 2001. *Announcing the advanced encryption standard (AES)*, Federal Inf. Processing Standards Pub., Vol. 197.

Dusart P., Letourneux G., Vivolo O. 2003. *Differential Fault Analysis on A.E.S,* E-print: 010.

Piret G., Quisquater J-J. 2003. *A Differential Fault Attack Technique Against SPN Structures, with Application to the AES and KHAZAD*. In the proceedings of the 5th international conference on Cryptographic hardware and embedded systems, LNCS 2779.

Schmidt J., Hutter M. 2007. *Optical and EM Fault-Attacks on CRT-based RSA: Concrete Results.* Ed. Austrochip 2007*,* 15th Austrian Workhop on Microelectronics, 11 October 2007, Graz, Austria, Proceedings.

Schmidt J., Kim C.H. 2009. *Information Security Applications*. Chung K., Sohn K., Yung M. eds., Berlin, Heidelberg: Springer-Verlag*,* pp. 256-265 ISBN 978-3-642-00305-9. DOI 10.1007/978-3-642-00306-6–19.

Skorobogatov S.P. 2005. *Semi-Invasive Attacks – A New Approach to Hardware Security Analysis*. PhD thesis, University of Cambridge, Computer Laboratory.

Takahashi J., Fukunaga T. 2007. *Differential Fault Analysis on the AES Key Schedule*. E-print: 480.

Trichina E. 2003. *Combinational Logic Design for AES SubByte Transformation on Masked Data*. E-print: 236.

Tunstall M., Mukhopadhyay D., Ali S. 2011. *Differential fault analysis of the advanced encryption standard using a single fault.* In the Proceedings of the 5th IFIP WG 11.2 international conference on Information security theory and practice: security and privacy of mobile devices in wireless communication. Heraklion, Crete, Greece.

Wang F., Agrawal V.D. 2008. *Single Event Upset: An Embedded Tutorial.* Proc. of 21st International Conference on VLSI Design.

Yen C., Wu B. 2006. *Simple Error Detection Methods for Hardware Implementation of Advanced Encryption Standard.* IEEE Trans.Comput., jun, 55(6): 720–731 ISSN 0018-9340. DOI 10.1109/TC.2006.90.

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

MOVING TOWARDS IPV6

Aurel Machalek
University of Luxembourg, Luxembourg

Key words: Internet Protocol version 4, Internet Protocol version 6, Seamless communication, TETRA.

Abstract

This document describes the possibilities for Internet Protocol communications in crisis situations. Its main goal is to show IPv4 and IPv6 solutions developed during the lifetime of the SECRICOM project. Communications technologies in current use are showing their limitations. Whether reacting to a small incident or a great catastrophe, first responders increasingly need to share information such as video, images or other data. Society's evolving expectations concerning safety can be compared with the development of Internet protocols. We examine IP-based communication for crisis management, and show that it is ready to bind together currently fragmented technologies such as TETRA and analogue radios, providing a new dimension of interoperability, including cross-border communication.

ZMIERZAJĄC DO PROTOKOŁU IPV6

Aurel Machalek
University of Luxembourg, Luxembourg

Słowa kluczowe: IPv4, IPv6, „bezszwowa” komunikacja, TETRA.

Abstrakt

W artykule przedstawiono możliwości wykorzystania protokołu IP (w wersji 4 i 6) w zarządzaniu kryzysowym, ze szczególnym uwzględnieniem rozwiązań wytworzonych w ramach realizacji projektu SECRICOM. Zauważa się, że zbliżamy się do limitów możliwości obecnie wykorzystywanych technologii komunikacyjnych. Podczas akcji ratunkowych – zarówno niewielkich incydentów, jak i wielkich katastrof – u służb ratunkowych w coraz większym stopniu występuje zapotrzebowanie na wymianę informacji pod różnymi postaciami (np. wideo, zdjęcia oraz inne informacje). Zwiększające się oczekiwania społeczeństwa wobec aspektów związanych z bezpieczeństwem mogą być skonfrontowane z możliwościami rozwijającej się technologii. W artykule omówiono łączność w zarządzaniu kryzysowym opartą na protokole IP. Autor zauważa, że podejście takie jest już realizowalne oraz ma potencjał umożliwienia wymiany informacji między różnorodnymi i rozproszonymi sieciami (np. TETRA, systemami analogowymi), wprowadzając nowy wymiar interoperacyjności, także w komunikacji międzygraniczej.

Context and Motivation

The SECRICOM project addresses communication security, interoperability, and connection continuity. Its main deliverable is [...] *a system that ensures end-to-end secure transmission of data and services across heterogeneous infrastructures with real time detection and recovery capabilities against intrusion, malfunctions and failures* [...][1] A major aspect of the project's work was an assessment of the IPv6 protocol, and its advantages and disadvantages in crisis communications. This paper discusses the following topics:

– Research in to crisis management communication and IPv6;
– The exhaustion of the IPv4 address space;
– Motivation for applying an IP-based communication system;
– Solutions to common communications system failure scenarios in crisis situations;
– Benefits for end users.

Related work

Work on this topic started in 2006, with the launch of the U-2010 project: *U-2010 project overall objective is to provide the most capable means of communication and the most effective access to information to everybody required to act in case of accident, incident, catastrophe or crisis, while using existing or future telecommunication infrastructures. The U-2010 project will address the public safety issues by researching new emergency and crisis management solutions investigating on innovative and state-of-the-art communication technologies based on the current and new Internet technologies (i.e. Internet Protocol version 6)*[2]. This work has been continued in the SECRICOM project.

Problem statement

Our research addresses many of the security and interoperability issues that have been highlighted by catastrophes across the world. It must provide answers to these questions:

– Is the Internet ready to be used as an emergency communications channel?

[1] Copyright is held by the author/owner(s). D.O.W. SECRICOM Projekt.

[2] Copyright is held by the author/owner(s). D.O.W. U-2010 projekt.

– Can it provide seamless, secure and reliable communications?

– Can first responders use this technology?

– Is IPv6 suitable for use with the latest communications technologies, and can it solve these problems?

Research goals and methods

In order to deal with the increasing frequency and complexity of disasters, whether natural or man-made, forward-looking emergency agencies are developing IP-based architectures able to integrate innovative solutions for their evolving operational requirements. The end user adopting the SECRICOM framework will benefit from:

– Improved Situational Awareness: Real-time communications allow officers to make better decisions more quickly, resulting in safer communities and a more efficient public safety workforce.

– Network Reliability: Redundant wireless network connections ensure reliable communications while on the scene or in motion.

– Office network extended to the incident: Provides real-time access to remote broadband applications for first responders in the field.

– Confidentiality of Information: Ensures information shared between an Emergency Operations Centre and first responders is secure and available only for public safety use.

– Interoperability: Provides a standards-based network platform which enables communications interoperability.

Future network growth requires that Internet-enabled devices can be assigned, used and – most importantly – be reachable anywhere via a globally unique IP address. Without sufficient global IP address space, applications are forced to work with mechanisms that provide local addressing for local internal communications and workaround "fixes" to communicate externally across the Internet. While waiting for a permanent address space solution, there have been numerous optional fixes to try to overcome the address space limitations. These include Network Address Translation (NAT), Classless Inter-domain Routing (CIDR) and extensions to IPv4.

Network Address Translation (NAT) allows multiple devices to be hidden behind one or more real IPv4 addresses. Such mechanisms restrict the end-to-end transparency of the Internet. While NAT has to some extent delayed the pressure on IPv4 address space for the short term, it places severe restrictions on capabilities for bi-directional communication between application endpoints. While a client behind a NAT device can communicate out to servers on the Internet (the client-server communication model), that same

Table 1

IPv6 Feature	Advantage (Compared to IPv4)
128-bit Addressing [RFC 2460]	Scalability from 2^{32} potential addresses to 2^{128} addresses, vastly expanding usable unicast and multicast address space
End-to-End Addressing [RFC 2460]	Reintroduces the end-to-end model to greatly lower the cost and complexity of peer-to-peer communications by eliminating the need for Network Address Translation (NAT)
Network Layer IPsec [RFC 2460, 4301, others]	Improved security support via IP layer security (IPsec) making it cheaper to deploy VPN-like security for all applications
New QOS Support [RFC 2460]	Potential new QOS capability through use of IPv6 flow labels
Auto configuration [RFC 2461, 2462, others]	Improved "plug and play" support using IPv6 link-local addressing, scoped multicasting & anycast support to automatically self-configure and discover neighbour nodes, routers, and servers
New Address Types [RFC 4291, 4193]	New addressing options for link local, anycast, intra-domain3, and globally unique Internet communications.
Security Addressing [RFC 3041, 3972]	New security addressing options for randomly-generated addresses to protect privacy, and cryptographically-generated addresses used to sign and authenticate messages
Enhanced Multicast Features [RFC 3306, 3956, 4291]	Enhanced local and global multicasting support scoped multicasting, and a tremendous expansion of usable multicast address space. Each site receiving an IPv6 prefix can generate 2^{32} globally routable multicast groups[1]. IPv6 multicasting can support creation of new geospatial and community-of-interest information distribution paradigms. Embedded-RP removes the need for IPv4 MSDP, simplifying deployment. Multicasting is a key feature used extensively for IPv6 autoconfiguration features
Multihoming Features [RFC 4291]	Multiple addresses can be assigned to IPv6 network interfaces. Use of different addresses can be used to differentiate link-local, intra-domain, and global messages. Addresses can be assigned and utilized for specific security, reliability, load-balancing, and QOS policies.
Simplified Header [RFC 2460]	Improved header structure that retains only the absolutely necessary header fields and eliminates IPv4's unnecessary CRC checksum fields. Speeds up packet processing in routers and makes basic IPv6 header more compressible than IPv4 for low data rate wireless and dial-up connections.
Extensible Headers [RFC 2460]	Extension headers are an extremely powerful feature that allows additional protocol-level information to be added to the basic IPv6 header. This allows additional protocols and services such as IPsec and mobile IPv6 to easily be integrated on top of the basic IPv6 protocol
Advanced Network Services [RFC 2460, 3775]	Basic IPv6 features and extension headers can be leveraged to build more powerful network services for mobility, security, QOS, peer-to-peer applications, etc. Mobile IPv6 improves on IP mobility for IPv4.

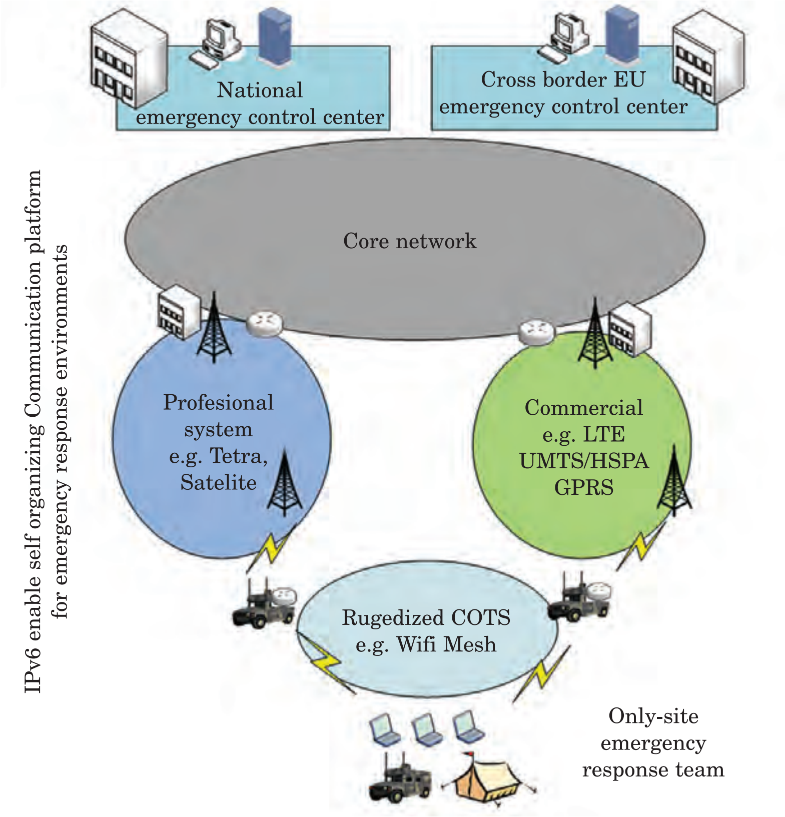


Fig. 1. High-level system overview

client cannot be guaranteed to be accessible when external devices wish to establish a connection to the client (as typified by the peer-to-peer communication model). NAT breaks the end-to-end principle of the Internet, restricting many applications that could be deployed as peer-to-peer to being deployed within a more complicated and expensive modified client-server model that relies on communications gateways and intermediate servers to connect hosts. NAT inhibits the evolution of next-generation applications that demand IP address space and direct remote connectivity into business premises and home networks (e.g. from IP-enabled mobile handsets). IPv6 reintroduces the ability to provide true end-to-end security that is not always readily available through a NAT-based network (Ipv6 Forum Roadmap...).

IPv6 has numerous technical features which, when compared to IPv4, make it a more powerful and flexible framework to deploy next-generation network applications and services (tab. 1).

Some of the operational benefits brought by adopting the SECRICOM system are:

– Extension of emergency mission network to Outdoor & Mobile Environments.

– Seamless mobility & continuous access.

– Network Security, Scalability & Manageability.

– Standards-based interoperable solution with investment protection.

Conclusion

The technical benefits of IP are:

– Redundancy: Redundant network connections across multiple wireless networks using standards based mobile IP.

– Security: Secure connectivity to the vehicle using network encryption, firewall, intrusion detection, FIPS 140-2 compliance.

– Connectivity: Mobile router provides network connectivity for wired or wireless client devices in or around a vehicle Network Management: Use the same network management tools to manage the office network and mobile network.

– Multi-Media Applications: Broadband wireless network for voice, video, and data applications for real-time communications with mobile network.

– Wireless Agnostic: Interfaces with 802.11b/g, licensed 4.9GHz, 3G and future wireless networks.

– Modular Design: Enclosure slots allows module expansion (e.g. video).

Translated by Authors

Accepted for print 30.06.2012

References

U-2010 project: http://www.u2010.eu.

Next Generation Public Safety Communication Networks and Technologies (NgenSafe'09).

Seinhardt G. *Blackblot Procedural Requirements Management Model*. Rev. 2.1. Available at: http://www.blackblot.com/prm-model/.

Wikipedia, *Emergency management*. Available at: http://en.wikipedia.org/wiki/Emergency_management.

IPv6 Forum Roadmap – http://www.ipv6forum.com/dl/forum/wwc_ipv6forum_roadmap_vision_2010.pd.

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

SECRICOM SILENTEL – SECURE COMMUNICATION INFRASTRUCTURE FOR CRISIS MANAGEMENT

Vladimir Hudek, Miroslav Konecny, Stefan Vanya

Ardaco a.s., Bratislava, Slovakia

Key words: crisis management, Secure Docking Module (SDM), Secure Agent Infrastructure (SAI), SECRICOM, first responder, push to talk, PTT, seamless communication, end-to-end security.

Abstract

The major incidents such as terrorist attacks in Madrid, London or huge natural disasters (fires, floods etc.) showed that the factor that inhibits the full performance of emergency response is the lack of interoperability between communication systems of different responders over Europe. The SECRICOM project built a solution that seamlessly integrates the communication systems of different organizations located in different EU member states. The term "seamless" means that the communicating parties may use their own communication devices or can use new Secricom Silentel enabled devices over the SECRICOM infrastructure without worrying about what technology or communication system is used by other users. The provided solution offers high level of security in a cost efficient way since it reuses the existing communication infrastructures including public and dedicated ones (eg.: Tetra Radio, GSM, UMTS, Internet).

SECRICOM SILENTEL – BEZPIECZNA PLATFORMA KOMUNIKACJI DO ZARZĄDZANIA KRYZYSOWEGO

Vladimir Hudek, Miroslav Konecny, Stefan Vanya

Ardaco a.s., Bratislava, Slovakia

Słowa kluczowe: zarządzanie kryzysowe, Secure Docking Module (SDM), Bezpieczna Infrastruktura Agentowa (SIA), SECRICOM, responder, PTT opcja naciśnij i mów, „bezszwowa" komunikacja, bezpieczeństwo połączeń.

Abstrakt

Podczas poważniejszych sytuacji kryzysowych, jak ataki terrorystyczne w Madrycie czy Londynie, a także klęski żywiołowe (pożary, powodzie etc.), okazało się, że głównym czynnikiem ograniczającym wydajność reagowania służb ratunkowych jest brak interoperacyjności między różnymi systemami łączności. W ramach projektu SECRICOM zbudowano rozwiązanie, które umożliwia „bezszwową” integrację środków łączności wykorzystywanych przez różne służby ratunkowe (także działające w różnych krajach). Pod pojęciem „bezszwowa” jest rozumiana możliwość wykorzystywania przez służby ich własnego sprzętu komunikacyjnego (bądź urządzeń wyposażonych w SECRICOM Silentel) na podstawie infrastruktury SECRICOM bez konieczności brania pod uwagę typu sprzętu używanego przez inne podmioty. Zaproponowane w ramach projektu rozwiązanie zapewnia wysoki poziom bezpieczeństwa oraz efektywność kosztową wynikającą z wykorzystania obecnie istniejących rozwiązań telekomunikacyjnych (zarówno publicznych sieci, np. GSM, UMT, Internet, jak i dedykowanych rozwiązań, np. TETRA).

SECRICOM Project Context

The major incidents such as terrorist attacks in Madrid, London or huge natural disasters (forest fires, floods etc.) showed that the factor that inhibits the full performance of emergency response is the lack of interoperability between communication systems of different responders over Europe. The complexity of this problem is increased by the number and structure of responder organizations involved simultaneously at different types of emergency events. These organizations have absolutely different hierarchical governance structures in place, maintain different cultures, pursue different goals and also use different means and tools for communication. After all, the cooperation of these stakeholders is essential for the successful resolution of the most of crisis situations.

The SECRICOM project has built a solution that seamlessly integrates the communication systems of different organizations located in different EU member states. The term “seamless” means that the communicating parties may use their own communication devices or can use new Secricom Silentel enabled devices over the SECRICOM infrastructure without worrying about what technology or communication system is used by other users. Moreover the new capabilities and the secure communication itself can be extended to mass market devices as PCs, mobile phones and tablets to increase cost effectiveness and comfort of users. Of course the preservation of life is the top priority but as the incident evolves the need for the confidentiality and reliability of information is getting more and more important. In SECRICOM, the communication sessions are ciphered, thus end to end security assurance is provided by the SECRICOM enabled devices.

Since most of potential hacker attacks are targeting endpoints in the security infrastructures, the SECRICOM solution introduced comprehensive network monitoring, intrusion detection, a hardware chip called Secure Docking Module and Trusted Docking Station that verifies the integrity of endpoint over the complete chain of trust starting from hardware itself. Thus the SECRICOM solution can provide a high level of security in a cost efficient way since it reuses the existing communication infrastructures including public and dedicated ones (Tetra Radio, GSM, UMTS, Internet, etc.). The use of different means of communication in parallel increases the availability of communications and makes the SECRICOM solution resilient.

During the analytical phase of the project different scenarios were identified with the involvement of end user group and from different points of view at different levels in the organizational hierarchy. The project results were demonstrated in incremental way as more and more features and capabilities were added during several shows and civil protection exercises. The SECRICOM project has proven the feasibility of cooperation among emergency responders in the face of many of constraints.

Operation Requirements of Communication System

Crisis management deals with unpredictable catastrophic events including terrorism (e.g. Madrid and London, in the future CBRN) and crime, natural disasters (including pandemics, earthquake and hurricane that are exacerbated in poor countries) and major industrial accidents/technological disasters (Toulouse 2001, tanker Prestige, Hertfordshire fire etc.). The increasing heterogeneity of potential crisis situations led to (and will lead to) establishment of rescue organisations accountable for new types of threats. The increasing number of organisations with different command structure and communication systems created the need for a seamlessly integrated and reconfigurable communications infrastructure for use inside and outside the EU. Moreover the future communication systems are required to be secure, smart, open, restorable and ubiquitous. The main challenge of the SECRICOM project was to **exploit the existing publicly available communication network infrastructure with the possibility to add more sophisticated tools**.

Interoperability between various responder agencies may be defined as the capability of two organizations or discrete parts of the same organization to exchange decision-critical information and to use the information that has been exchanged.

The Figure 1 beside depicts the interoperability stack in which compatible high level objectives and physical interoperability (electrical/mechanical interconnections and transmission signalling) reside at the top and bottom, respectively. Lack of interoperability has been recognized as a significant reason for lack of effectiveness of a given crisis.

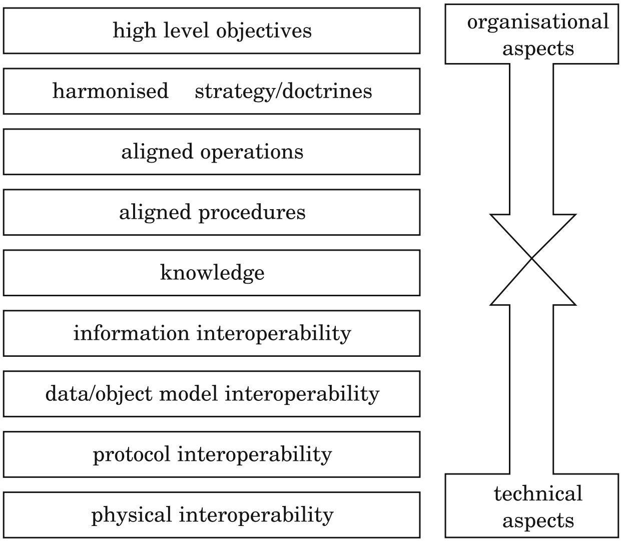

high level objectives
harmonised strategy/doctrines
aligned operations
aligned procedures
knowledge
information interoperability
data/object model interoperability
protocol interoperability
physical interoperability
organisational aspects
technical aspects

Fig. 1. Interoperability aspects

This is due to the consequential weakening of the process of coordination between emergency response personnel during the crisis management which is brought about by the resulting increased likelihood of poor critical-decision making.

SECRICOM Silentel Architecture

Secricom Silentel is a client-server communication system using Internet Protocol (IP) as illustrated in the Figure 2. It enables the use of mobile devices (i.e. smartphones, tablets or computers) together with technologies currently deployed (i.e. Tetra, SDR, etc.) for daily routines and crisis communication of public safety services. It optimizes and protects the way teams of people communicate without being concerned about misuse of information.

The main parts of the SECRICOM Silentel architecture are as follows:

– The Communication Server is a secure switching center module interconnecting all users of the system – as described by Figure 3.

– The Certification Authority is the trust module for Server and Users certification creation, validation and revocation (user can use his own CA as well).

– The Operator Studio serves as a tool for user account management and their personal contact list definition.

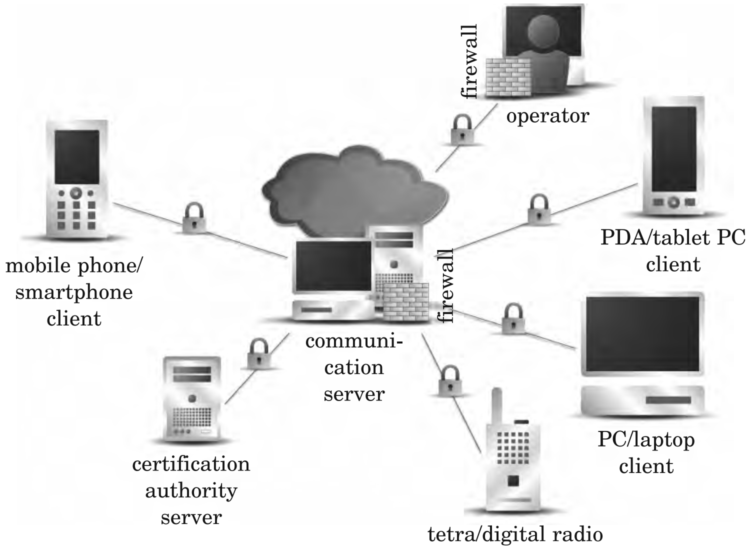


Fig. 2. Secricom Silentel (PTT) high level architecture

The SECRICOM Silentel client application communicates only with the server. This approach was influenced by the fact that there are no static IP addresses and there is no support of IPv6 by current GSM network operators. There are two main communication channels; one is the signalization protocol (SIP) and the other is RTP for audio transmission. SIP uses TCP and RTP uses the UDP protocol.

During the session, the server only uses the SIP protocol to send updates about the presence status of user's contact list; there are also regular "still-alive pings" from client to server.

As soon as the user starts a session, SIP is used to transmit more information, such as Talk Burst requests, text messages, pictures, information about users in the same session (users come and go) and encryption keys. The information structures transmitted by SIP are encoded in ASN.1. Each session is encrypted by different AES256 key. The same AES key is used for encryption of voice. When the user requests the Talk Burst, he can receive it back from the server – the application plays a signalization tone and changes Press-To-Talk button's colour. The audio recording is transformed by AMR compression, then the AES encryption takes place and RTP transmission is started. The server knows what users are in what session and simply serves as router; the server does not process the audio information in any way.

Security provisions of SECRICOM Silentel include besides ciphering functions also user authentication and management tools for the control of user permissions. Unique user name and password with physical gridcard authentification protects user accounts. Smart card (microSD) and PIN (electronic signature) support the security of access. The Operator studio enables to manage rights, groups and visibility of clients with a possibility to block a suspiciously behaving account in real time.

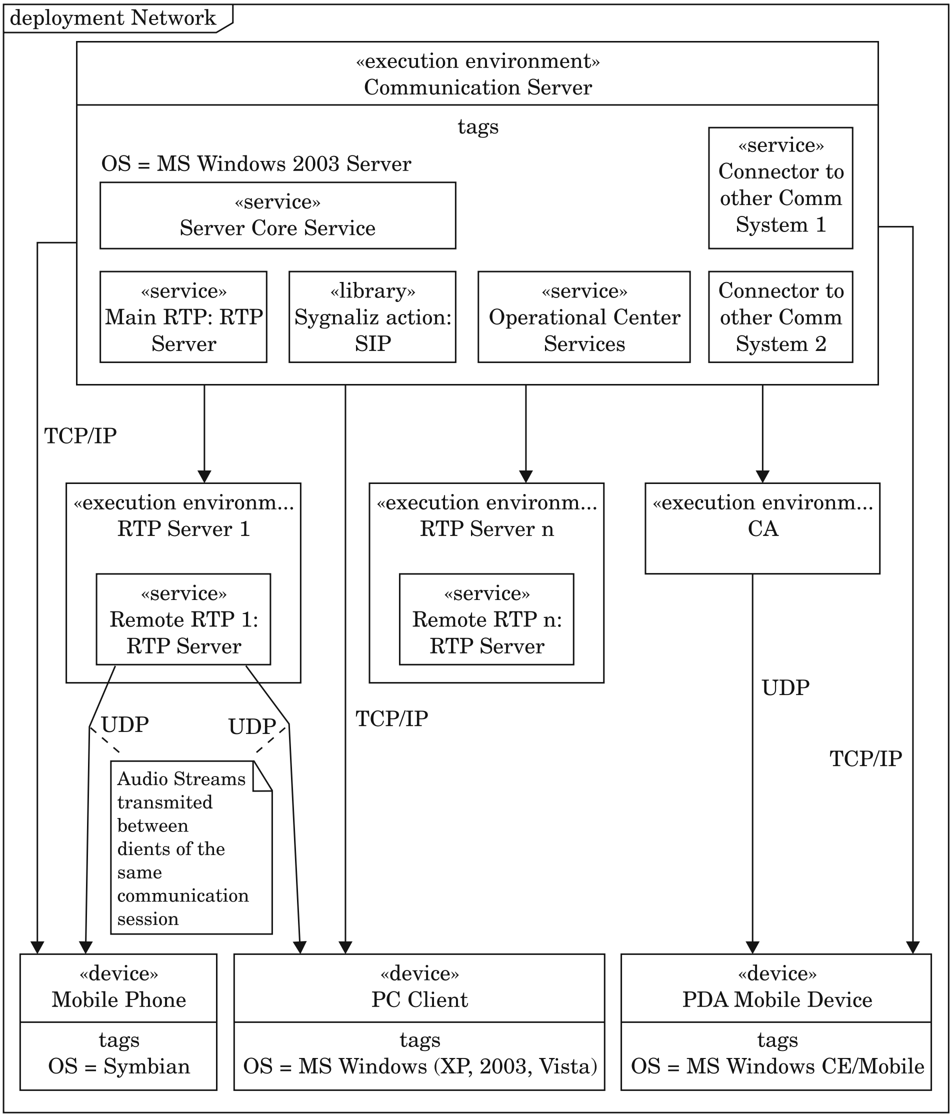


Fig. 3. SECRICOM Silentel Server Components

The client and server have a layered architecture. Each layer, process or module communicates with the rest of the system using asynchronous messages. The concept was proven on various supported operating systems of Secricom Silentel client, i.e. Symbian, Windows Mobile, Android, iOS as well as Windows. Communication gateway was implemented for Tetra and CB radio services, and the implementation allows also integration with further systems. The concept is described on Figure 4.

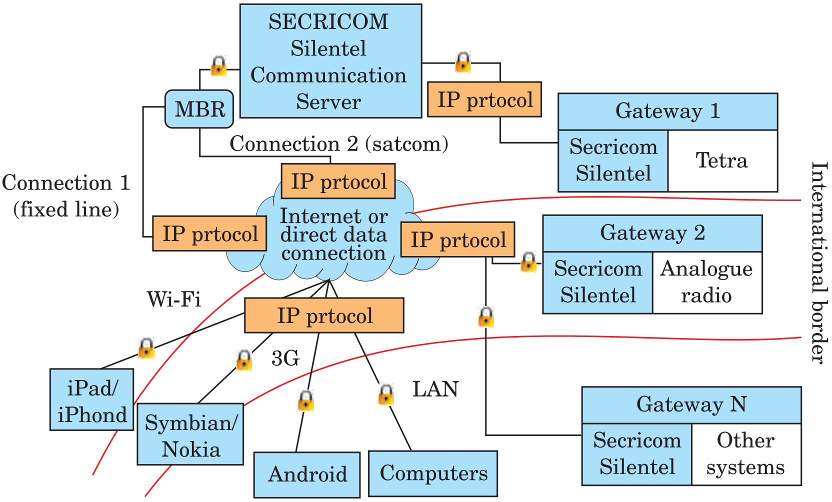


Fig. 4. SECRICOM Silentel Integration with Tetra and Analogue Radio

SECRICOM Silentel Capabilities

SECRICOM Silentel application features were designed to support the operation of public safety agencies and other actors in PTT

daily routines and crisis situations. As defined by the user requirements analysis undertaken by SECRICOM consortium, **voice conversation** still remains the very first requirement. SECRICOM Silentel enables to build a one-to-one call with full duplex voice transfer or one-to-many group call (Fig. 5) with half duplex voice (press-to-talk). The management of any group is flexible with a possibility to add additional members by any of users involved in conversation. This voice service can be used for seamless involvement of actors using different devices and located in different countries.

Another feature offered by SECRICOM Silentel is **textual service** such as instant messaging (one-to-one or one-to-many) or long message up to 1000

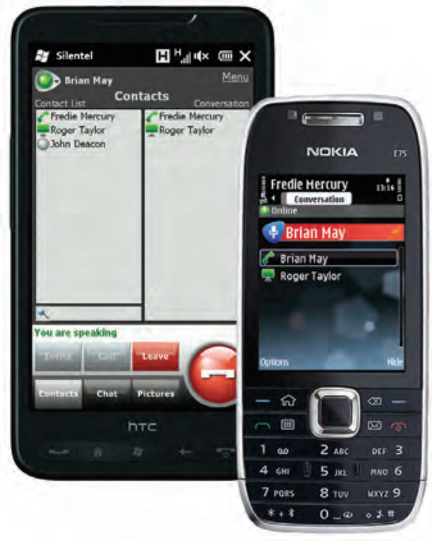
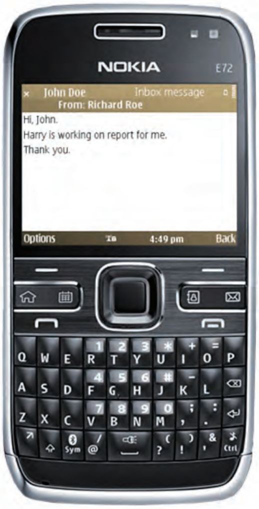


Fig. 5. Secricom Silentel Voice Conversation

unicode characters (Fig. 6). These were proven by a demonstration exercise as an useful tool for correct spelling of name (instant messaging) or treatment instructions for hazardous chemical sent by external expert (long message). Both services enable audit trail within protected environment and also delivery and read receipt.



From: Richard Roe

Hi, John.
Harry is working on report for me.
Thank you.

Fig. 6. Secricom Silentel Texting

Emergency response requires all senses in action including visual sense. Visual sense can be supported by **use of images** in daily business and special occasions. SECRICOM Silentel enables to send images in conversation groups, both saved from device memory and taken in real time. This feature was used for sending an image of hazardous barrel label in Portsmouth, UK to an external expert located in Poznan, PL. Similarly, some devices can use hand drawing over defined canvas to share directions – like sketching directions over building floor plan during evacuation (Fig. 7).

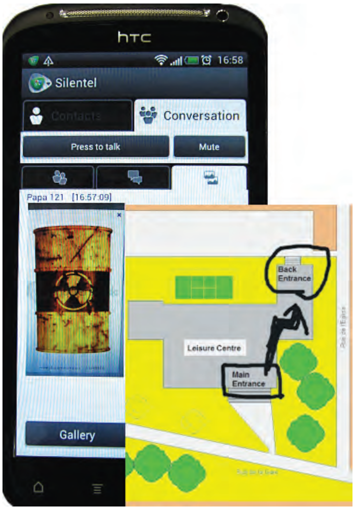


Fig. 7. Secricom Silentel Images Transfer

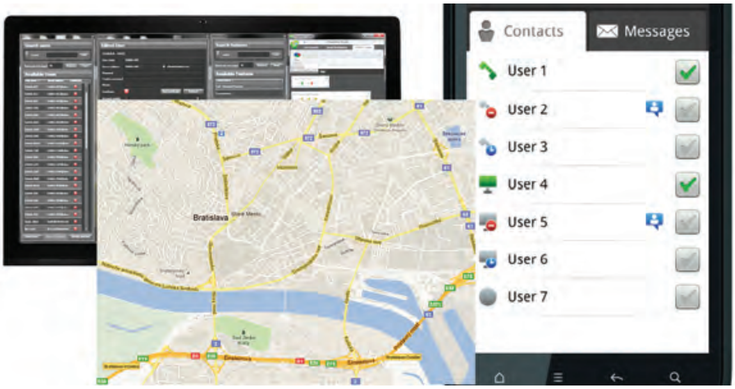


Fig. 8. Secricom Silentel Operator Studio and GPS positioning

When managing a group of actors, their location and availability really matters. SECRICOM Silentel has a **GPS positioning** feature, so a master user (such as an operator) can see the current location of different resources on a map in real time. Every installed application contains a secured contact list with presence status – users can simply select one of pre-defined statuses, such as available/emergency state/non available/etc. This allows identifying **available resources** at a glance (Fig. 8).

Conclusions

SECRICOM Silentel is a scalable ICT solution consisting of a communication server, a certification authority, an operator studio, end user application and interfaces to other systems. It was designed and developed in line with the SECRICOM vision – to provide technologies supporting performance of emergency responders in a comfortable way on standard devices and networks. It introduces new features for standard devices such as smartphones and laptops – they are turned into powerful and secure communication tools suitable for current actors and allowing involvement of new actors with certain responsibilities into emergency actions. These can include officers in emergency agencies, local government officials (county, town) and also infrastructure services (water/gas/electricity/transport).

SECRICOM Silentel is open for further integration with systems currently in use by public safety services. Its ambition is not to replace, but supplement them by adding new devices and actors to emergency teams. Concerns about connectivity of IP based systems can be answered by listing priority users in public/private networks, extendable network means and/or use of Multi-Bearer-Router for seamless connectivity.

SECRICOM Silentel supports heterogeneous actors cooperating during emergency response, both cross-border and inter-agency. Security is taken as an integral part of the service in sense of transfer of data, access and user management. It allows using the infrastructure for security-sensitive operations as well. These features were demonstrated in an integrated demonstration of the project and are to be developed into a product by ARDACO in following months.

Translated by Authors

Accepted for print 30.06.2012

References

EUROPOLTECH 2011. International Fair of Technology and Equipment for the Police and National Security Services. Fair Magazine Article http://www.energia.sk/clanok/veda-a-vyskum/secricom-predstavil-vysledky-dvojrocneho-vyskumu/1885/

http://www.itti.com.pl/en/eu-projects/on-going-projects2.html

http://www.secricom.eu/

http://www.secricom.eu/images/articles/SECRICOM-HN-29-7-2010.pdf

http://www.secricom.eu/images/articles/Secricom-leaflet-4-2010.pdf

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

DELICATED TO THE CRISIS MANAGEMENT SOLUTION COMMUNICATION SECURITY MONITORING AND CONTROL CENTRE

Saioa Ros, Oscar López, Mikel Uriarte

NEXTEL S.A., Zamudio, Spain

Key words: Communication Infrastructure, Risk Assessment, Security Mechanisms, Security Model, Awareness, Control.

Abstract

An emergency situation often occurs as a result of unpredictable events, and as a consequence, existing communications may either get collapsed or congested. The aim of the Communication Security Monitoring and Control Centre (CSMCC) solution proposed in SECRICOM is to provide a suitable security framework that enables the development of security service. These services give response to individuals and institutions operating in heterogeneous communication infrastructures, when responding to major incidents. In the light of this objective, a Security Model has been designed facilitating the measurement of operators' and end customers' confidence in the security of the communication infrastructure, and addressing security challenges in terms of a distributed and heterogeneous solution. The proposed Security Model has been supported by the Security Middleware Service and Framework, which is responsible for measuring, documenting and maintaining the security level of the services provided by the SECRICOM communication system.

ROZWIĄZANIE COMMUNICATION SECURITY MONITORING AND CONTROL CENTRE PRZEZNACZONE DO ZARZĄDZANIA KRYZYSOWEGO

Saioa Ros, Oscar López, Mikel Uriarte

NEXTEL S.A., Zamudio, Spain

Słowa kluczowe: infrastruktura komunikacyjna, ocena ryzyka, mechanizmy bezpieczeństwa, model bezpieczeństwa, świadomość sytuacyjna, sterowanie.

Abstrakt

Sytuacje kryzysowe występują zazwyczaj w wyniku nieprzewidzianych wydarzeń, co implikuje problemy z niedostępnością lub przeciążeniem systemu łączności. Zadaniem zaproponowanego w ramach projektu SECRICOM rozwiązania Communication Security Monitoring and Control

Centre (CSMCC) jest bezpieczeństwo usług z nim związanych. Usługi te są kierowane do podmiotów obsługujących różnorodne infrastruktury telekomunikacyjne na potrzeby zarządzania kryzysowego. W tym celu opracowano model bezpieczeństwa ułatwiający pomiar bezpieczeństwa infrastruktury komunikacyjnej oraz wspierający rozwiązywanie problemów wynikających z heterogenicznych i rozproszonych rozwiązań. Zaproponowany model jest wspierany przez rozwiązanie Security Middleware Service and Framework, który jest odpowiedzialny za mierzenie, dokumentację oraz utrzymanie poziomu bezpieczeństwa usług oferowanych przez system SECRICOM.

Introduction

The aim of the communication security control centre is the design of a Security Model for the SECRICOM communication infrastructure suitable for secure and interoperable communications under crisis situation. Additionally it has been supported the development of a Security Middleware Services and Framework to measure, document and maintain the security assurance level of services based on telecommunication systems.

On the technical aspect, SECRICOM presents important research challenges for the design of information security management solutions. SECRICOM has provided a heterogeneous communication infrastructure independent from the different civil forces that participate in crisis incidents, and also a boarder cross distributed interoperable solution that allows cooperation among agencies from different countries. The communication infrastructure proposed in the project is able to be dynamically reconfigurable in terms of security settings to address diverse communication-data exchange contexts. Last but not least, SECRICOM solution has provided seamless operation for end-users, so that a user can make use of the communication infrastructure in a transparent manner to the underlying security mechanisms.

From a more pragmatic overview, the challenge of engaging with end-users has been covered, in such a way that meaningful security requirements has been defined for the communication infrastructure. The first responders and commanders are under stress and very busy while trying to handle the situation. It is not likely that they turn to different communication system than the one they are used to from day-to-day operations just because of security issues. The system is prepared to be used in daily and able to handle the emergency situation specifics. The security mechanisms and policies have been built in by default. As a result of this, SECRICOM provides secure communication services for day-to-day operations and on-site-deployable infrastructure for crisis situations in a very efficient manner.

Security Objectives

These challenges are met in the design of a Security Model suitable for secure and interoperable communications under crisis. The scope of the Security Model is based on the structure of security objectives (Information Security) that represent the principles on which an effective security has to be established.

By ensuring confidentiality any unauthorized disclosure of communications between two or more parties is prevented. By ensuring integrity data cannot be manipulated during the transmission. Indeed, integrity guarantees that the recipient of some data will realize if any alteration of the originator's message has been done. Additionally, integrity of the data includes the authentication of the user source, which guarantees that network entities are not pretenders. Lastly, by ensuring availability users are always sure that information and resources are available. Though it is not possible to completely avoid Denial-of-Service type of attacks, SECRICOM components such us network monitoring, network control, multiple bearers, policy based routing and dynamic reconfiguration greatly improve the overall service availability and thus provide increased continuity of business processes during disaster relief operations.

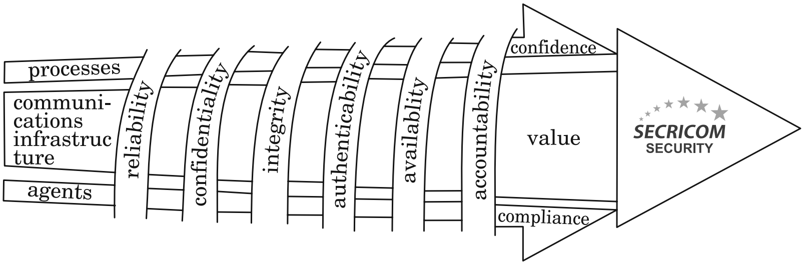


Fig. 1. Security Objectives

Availability of SECRICOM services is guaranteed not only through a balance of risk reduction measures but also through recovery options. The IT Service Continuity Management (IT SCM) plan for SECRICOM ensures that even in case of disaster the system will continue to provide the services. Provision of pre-determined and agreed level of IT services is particularly important for crisis management systems and even in case of interruption it has to support the minimum requirements, by means of the identification of required minimum level of SECRICOM services as well as the development of an IT risk mitigation program.

Responding also to project high level objectives, the design and implementation of assets must be done considering security as an integrated design element in order to ensure the availability, integrity and confidentiality of data and system resources supporting the key security functions.

Secricom Risk Assessment

Taking into account the complexity of the SECRICOM system, a single security technology has not been enough to overcome all the security objectives. The used method has consisted of a set the requirements for the security model that have to address all the security challenges. Thus, firstly a risk assessment (STONEBURNER 2002) and analysis of the heterogeneous communication infrastructure of SECRICOM has been done. The system architecture has been analysed in order to determine the value of assets in terms of impact or criticality for the whole system. SECRICOM infrastructure consists of Secure Agent Infrastructure (SAI), Push-To-Talk Service (PTT), IP Core Network, Legacy network gateways and Communication Monitoring and Control Centre. Furthermore, assets evaluation has been conducted, regarding the offered communication services as well as their relevance for the system.

SECRICOM critical networking, information and operational assets analysis has been helpful to identify the potential weaknesses (Common Weakness...) of the assets. It also permits to estimate the probability of exploiting particular vulnerabilities by a given threat and the potential impact that it may have on the operation of the system. Therefore, it is possible to determine the risk level of each asset. The outcome has been a risk treatment plan and a set of security requirements that need to be fulfilled in order to create an effective security model for a pervasive and trusted communication infrastructure.

In the aim of validating and updating drafted security requirements, a user team workshop was held on 13 April 2011 in London UK, in order to enhance SECRICOM requirements of the security model. The workshop aimed to obtain experience based statements, remarks and suggestions from individuals well versed in crisis management in a structured framework in order to establish the user requirements in terms of the security of the information carried on communications systems used by first responders during crisis management. The impact analysis was set against the background of Civil Protection Agency operational outcomes, such as the importance of having or not having a particular communication asset with reference to saving lives, ability to conduct emergency services, provision of local contingency services, impact on judicial proceedings and impact on foreign relations.

Information Exchange Requirements (IER) analysis results were presented as the starting point for this user team workshop in terms of considering the varying needs for security of communications assets depending on the differing three command levels, that is, strategic, tactical and operational levels. IER describes the process used to enable the Capability/Interoperability shortcomings of current Crisis Management communications systems to be identified. From this study it is possible to conclude that at strategic level, there is an identifiable need for data style communications to assist in decision making and the provision of strategic direction. At the operational level, which is mobile by nature, there is a high level of voice communications. Sitting between the operational and strategic level is the tactical level, which can be fixed and mobile but predominately nomadic (temporarily located at a location for a period of time with a capability to move). This level reflects the need for voice and data capability. What is irrefutable is that all levels have an increasing need for data capability. Thus, in the figure below it is shown that although Voice remains the most significant method of Information Exchange, Messaging (email / text) and other data types, such as web services (understood as services that will be consumed by final users via web interface), file transfer and video, also emerge as prominent, as shown in the figure below.

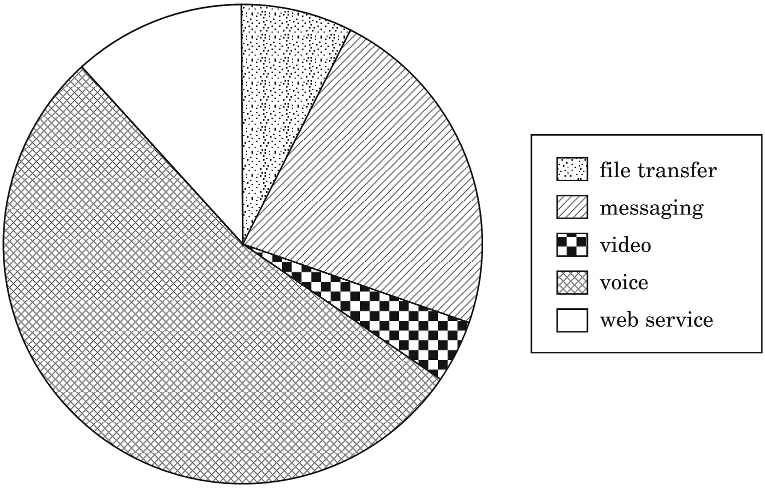


Fig. 2. IER distribution grouped by Communication Service

Back to the user workshop, these services have been the communication assets that have been analyzed and weighted in terms of impact against the operational outcomes described before and in terms as well of continuity and service operation for the typical information security components of Confidentiality, Integrity and Availability. A standard risk assessment process was adapted to assist the users in prioritising different security requirements for the communication assets. Subsequent analysis and evaluation of the workshop findings concluded in the following table that illustrates the impact perspective of all the three command levels for each communications asset for the core security requirements of Confidentiality, Integrity and Availability.

Table 1

Communication Asset value

Assets	Strategic			Tactical			Operational		
	confiden-tiality	inte-grity	availa-bility	confiden-tiality	inte-grity	availa-bility	confiden-tiality	inte-grity	availa-bility
Voice	5	5	5	6	5	5	6	6	6
Messaging (e-mail, data/text)	4	2	2	3	4	4	3	3	3
File Xfer (dokument)	4	2	2	3	4	4	3	3	3
Video	2	2	2	2	2	2	2	2	2
Web	0	2	2	3	1	1	3	1	1

Where “6” represents a “High” risk impact level and “0” represents “None”.

The key security requirements findings that emerged from the workshop were as follows:

– Voice communications at all three levels of command and between agencies, is seen as critical and therefore would require the highest level of security in terms confidentiality, integrity and availability.

– Messages and file transfer are seen as the next important communication assets in terms of security requirements.

– Web services are the least valued communication assets at all levels of command with video considered the next least valued communications asset.

– In comparison to Integrity and Availability, Confidentiality is considered a lesser requirement; probably reflecting the desire for as flexible communications as possible when life is at risk.

– Integrity, across all three command levels, is seen a key requirement (voice in particular) for all communications assets. I.e. a message sent in any medium during a crisis has to be received in a format that the receiver understands the message so that the requested action, decision etc is carried out.

– Availability of all communication assets, apart from voice which is viewed as essential, is seen as moderately important across all three levels of command with messaging and file transfer being viewed more important than video and web.

Hence, security objectives concentrate the security requirements that have to be covered by a set of appropriate countermeasures and security mechanisms (Implementing Network...) identified through the risk management process. These mechanisms support the three integral concepts of a holistic security program: detection, reaction and protection. Detection monitors for potential breakdowns in protective mechanisms that could result in security breaches. Reaction responds to detect breaches to thwart attacks before damage can be done. Protection provides countermeasures such as security policies, procedures, and technical controls to defend against attacks on the assets being protected. A security policy is a set of rules that dictate how sensitive data has to be managed, protected, and distributed. It provides the security goals that the system must accomplish. The level of security that a system provides depends upon how well it enforces the security policy, and a security model is a statement that outlines the requirements necessary to properly support and implement a certain security policy.

Secricom Security Model

The output of the risk analysis and the identification of the security mechanisms solutions that will provide the required security services is the set of security requirements presented in Table 2.

All these requirements are aggregated in the specification of a suitable security model. The SECRICOM Security Model proposes the application of these security principles to define the guidelines and rules for achieving a secure infrastructure. The security model does not define how to build or implement a secure infrastructure, but instead defines the properties, capabilities, processes and controls that a secure infrastructure contains to protect against a range of threats.

The SECRICOM Security Model defines two fundamental security objectives that are total visibility and complete control. Total visibility consists of identifying and classifying users, traffic, applications, protocols and usage behaviour; monitoring and recording activity and patterns; collecting and correlating data from multiple sources to identify trends and system-wide events; detecting and identifying anomalous traffic and threats. Complete control consists of hardening IT infrastructure, including individual devices and increasing network resilience; limiting access and usage per user, protocol,

Table 2

Security Requirements

Security Architectural Principle	Description
Defence-in-Depth	Never assume that a single control can provide sufficient risk mitigation for specific threat. Deploy multiple layers of controls to prevent, identify, and delay attacks in order to contain and minimize damage while an organization responds.
Service Availability and Resiliency	Ensure service availability through device hardening and by strengthening the resiliency of the network to adjust to and recover from abnormal circumstances.
Segregation and Modularity	Infrastructure is organized in functional blocks with distinct roles facilitating management, deployment, and securing of the devices and business assets within each block.
Regulatory Compliance and Industry Standards	Follow industry standards and best practices to facilitate the achievement of regulatory compliance.
Operational Efficiency	Simple and efficient configuration, deployment, and management of the infrastructure, throughout it entire life cycle, increase control and visibility allowing for faster auditing, troubleshooting, problem isolation, and incident response.
Confidentiality, Integrity and Availability	Security controls work to provide acceptable levels of confidentiality, integrity and availability of data.
Auditable and Measurable Controls	Security controls must be auditable and measurable to be effective.
System-wide Collaboration and Correlation	Infrastructure security is not a set of independent point solutions. Effective security requires sharing, analysis, and correlation of information from all system-wide sources.

service and application; isolating users, services and applications; protecting against known threats and exploits; dynamically reacting to anomalous events. The success of a security architecture and infrastructure implementation ultimately depends on the degree to which they enhance visibility and control. Without visibility there is no control and without control there is no security.

Secricom Security Model middleware

As a result of the performed analysis, the need of a SECRICOM Security Model software approach has been identified, which conducts, aggregates and integrates the necessary security protocols and mechanisms. This approach consists of an intelligent security framework and middleware service for adaptive information security management. It is represented by the Communication Security Monitoring and Control Centre (CSMCC). The CSMCC surveys if network infrastructures are resilient to both well-known and new forms of attack, implementing a cyclic assessment approach by means of network enumeration, network scanning and security assessment for dynamically plugged assets. CSMCC covers an increased scope for asset protection, including information protection mechanisms, access control and usage policies, scalable architecture, auditing tools and security assurance monitoring. It also manages improved detection processes and network forensic solutions in terms of new traffic patterns and enhanced event correlation mechanisms. Not only awareness and detection capabilities, but CSMCC has also enhanced reaction capabilities for hostile environments, such as traffic blocking, alternative routing solutions and isolation mechanisms. CSMCC includes fast recovery plans for crisis critical communications as well, to provide quick and efficient recovery mechanisms (CA eTrust...).

In order to be able to cover in an efficient manner the wide and heterogeneous communication infrastructure of SECRICOM, CSMCC follows a distributed architecture composed by a centralized server, which manages all the security information, and distributed sensors and agents, which are responsible for security information gathering.

Figure 3 describes the distributed architecture for security management, which enables to deploy the different components for monitoring and control, where some valuable enhancements with regards to legacy systems have been done.

– Agent level: Adaptive agents that fetch the state and the behaviour of the components of the SECRICOM multibearer and interoperable communication solution specifically deployed for this context.

– Sensor level: Adaptation and normalization of data specifically for each type of event generated by the agents.

– Correlation level: Intensive stress tests have provided a balanced performance between sensibility and time reaction, reducing the false positives events.

– Presentation level: Enhanced final user experience for modelling, configuration and reporting, dealing with alarm behaviour and treatment.

Indeed, it should be highlighted that one of the main strengths and unique features of the CSMCC platform in SECRICOM is the set of custom agents that have been deployed along the communication infrastructure in order to adapt to the security requirements that such a scenario as major crisis communication management needs. These enhanced agents provide new detection and action capabilities, such as adaptive routing features in case of network failure or congestion and VoIP traffic monitoring. Furthermore, in order to process the security events detected by this set of agents, the CSMCC platform has defined a set of patterns and policies adapted to the prioritization and correlation of these events, providing them of a meaningful context.

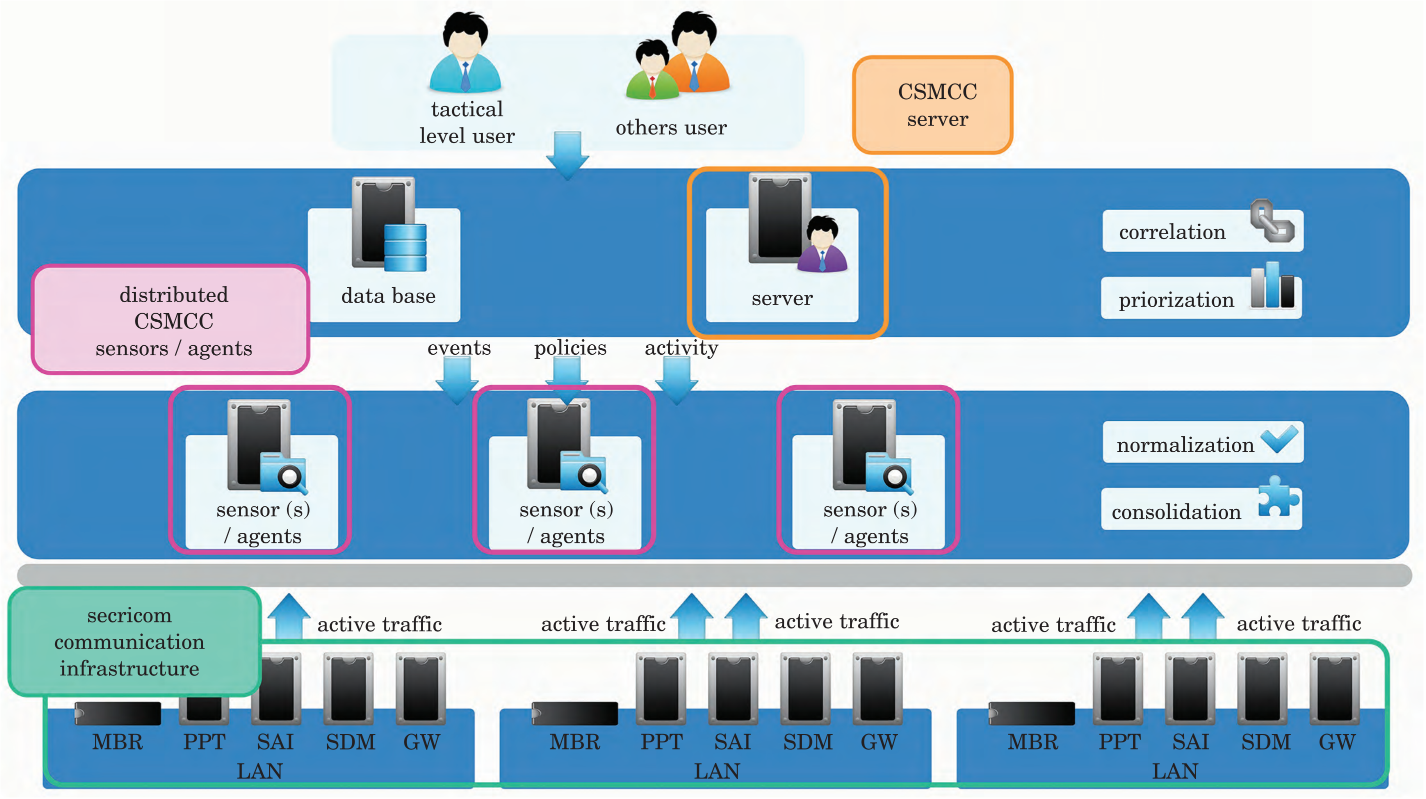


Fig. 3. Security Management Architecture

Finally, in order to cover the security objectives of total visibility and complete control of the security model, CSMCC capabilities are described below and they are split mainly in two categories, those related to security status monitoring and those capabilities related to reaction, that is, awareness and control capabilities.

- Security awareness:
  - Inventory and auto-discovery of network assets
  - Repository of hosts
  - Policy management to modify the importance of detected events
  - Sensibility policies and correlation management
  - Anomalous traffic detection
  - Intrusion detection
  - Event management
  - Alarm generation
  - Vulnerability discovery that allows a soon awareness of potential weaknesses
  - Network monitoring in terms of network usage and network latency
  - Host and service detail monitoring in terms of availability status
- Control services:
  - Traffic blocking
  - Traffic isolation
  - Alternative routing
  - Agent trust renewal
  - Configuration management

Results and Achievements

The main results achieved during the SECRICOM project in terms of communication security management can be summarized as follows:

– Identification of the security requirements to be covered by the SECRICOM Security Model.

– Delivery of the SERCICOM Security Model that defines security specifications and framework for a trusted communication infrastructure.

– Definition of the software prototype that allows the evaluation of the SECRICOM Security Model: Communication Security Monitoring Communication Centre (CSMCC) Platform.

– Deployment of enhanced and custom agents, responsible for collecting the state of the SECRICOM components.

– Development of a set of functioning patterns and policies adapted to the SECRICOM context.

Conclusions

The followed roadmap to design the communication infrastructure security monitoring and control centre starts with a risk assessment of the SECRICOM system. It consists on a deep analysis of the operation of the system, in order to define the key assets, identify their security vulnerabilities and evaluate the impact in terms of risk. The outcome of the risk assessment is a set of security requirements that the SECRICOM security model fulfils to provide an effective security management. This has been backed by a user team workshop that updated and validated these security requirements. The analysis of the security requirements results in a bunch of countermeasures and security mechanisms to mitigate the risk level obtained and protect the SECRICOM systems against the exposure to security threats. Finally, all these security principles and guidelines are aggregated into the SECRICOM security model, in order to achieve a secure communication infrastructure in a continuous way, ready for the dynamicity of the communications.

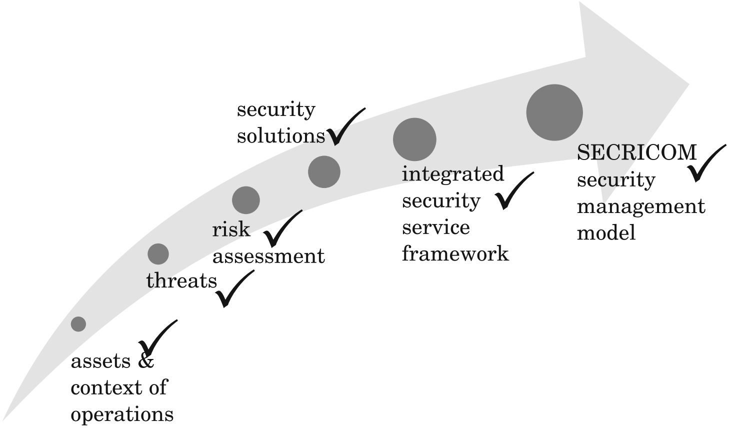

assets & context of operations ✓
threats ✓
risk assessment ✓
security solutions ✓
integrated security service framework ✓
SECRICOM security management model ✓

Fig. 4. Security roadmap

The security model is supported by a security middleware named Communication Security Monitoring and Control Centre (CSMCC) that provides not only security information collection and security status monitoring capabilities, but also active control mechanisms, providing enhanced protection, improved detection, faster reaction and stronger risk mitigation, more effective incident's impact mitigation and quicker restoration.

Translated by AUTHORS

Accepted for print 30.06.2012

References

Information Security. http://usdatasecurity.ch/itsecurity

STONEBURNER G., GOGUEN A., FERINGA A. 2002. *Risk Management Guide for Information Technology Systems*. Recommendations of the National Institute of Standards and Technology (NIST).

Common Weakness Enumeration, a Community-Developed Dictionary of Software Weakness Types; http://cwe.mitre.org/index.html

Implementing Network Security Mechanisms; CISCO; http://www.cisco.com/en/US/docs/ios_xr_w/iosxr_r3.3/security/design/guide/sg33impl.html

CA eTrust Security Information Management; "Imposing order on security information overload".

TECHNICAL SCIENCES
Abbrev.: Techn. Sc., No 15(1), Y 2012

ENHANCING SEAMLESS DATA TRANSFER WITHIN COMPLEX MESH ENVIRONMENTS WITH SECURE AGENTS

***Phil Entwisle*[1], *Steve Lawrence*[1], *Ondrej Habala*[2]**

[1] Qinetiq Ltd, Portsdown Technology Park, United Kingdom
[2] Institute of Informaties Slovak Academy of Sciences, Bratislava, Slovakia

Key words: Secure Agent Infrastructure (SAI), Multi Bearer Router (MBR), seamless data transfer, network optimization.

Abstract

QinetiQ's Multi-Bearer Router (MBR) has been designed to operate in harsh-dynamic environments and provide seamless data transfer within these situations. During certain network environments, specifically complex mesh, it has been witnessed that the MBR has not chosen the most appropriate interface quick enough to maintain seamless data transfer. This is mainly due to the MBR basing all of its decisions on local knowledge; hence a network change at two or more routers away can take too long to filter through the network for all MBRs to have up-to-date routing knowledge. Thus it was planned that a new MBR component would be created to assist in the distribution of all remote MBRs local knowledge to create a central global knowledge that can then be tailored to each MBR on the network. With updated remote knowledge the MBR can better support seamless data exchange in complex network environments without a large burden on the network or its constraints.

ZWIĘKSZENIE EFEKTYWNOŚCI „BEZSZWOWEJ" TRANSMISJI DANYCH W ZŁOŻONYM ŚRODOWISKU SIECIOWYM PRZEZ ZASTOSOWANIE SECURE AGENT

***Phil Entwisle*[1], *Steve Lawrence*[1], *Ondrej Habala*[2]**

[1] Qinetiq Ltd, Portsdown Technology Park, United Kingdom
[2] Institute of Informaties Slovak Academy of Sciences, Bratislava, Slovakia

Słowa kluczowe: Secure Agent Infrastructure (SAI), Multi Bearer Router (MBR), „bezszwowa" transmisja danych, optymalizacja ruchu sieciowego.

Abstrakt

Opracowane przez QinetiQ rozwiązanie Multi-Bearer Router (MBR) służy do zapewniania „bezszwowej” transmisji danych w szczególnie trudnym i zmiennym otoczeniu. Zaobserwowano, że w niektórych środowiskach sieciowych (w szczególności przy złożonych sieciach kratowych) MBR nie zawsze wybiera optymalne trasowanie. Głównym powodem takiego stanu rzeczy jest fakt, że decyzje w MBR są podejmowane na podstawie wiedzy dostępnej lokalnie. Zaobserwowano, że czas propagacji informacji bywa za długi, by cała sieć MBR miała aktualne informacje o dostępnych trasach. W następnej wersji komponentu MBR przewidziano wsparcie do budowy scentralizowanej i globalnej wiedzy. Byłaby ona dostępna dla poszczególnych urządzeń sieciowych, a te mogłyby wykorzystać i dostosować tę wiedzę do własnych potrzeb. Poszczególne routery MBR – dysponujące taką wiedzą – byłyby w stanie lepiej wspierać „bezszwową” wymianę danych w złożonych środowiskach sieciowych, z jednoczesnym minimalizowaniem obciążenia sieci.

Introduction

Initially the QinetiQ Multi-Bearer Routers (MBR) make all their decisions locally, occasionally it has to make assumptions whether or not the next MBR in-line has access to the required networks. Therefore, a need arises to allow MBRs to communicate and pass their local knowledge to all other MBRs for dissemination of the network routing information, especially with regard to meshed networks. After discussions and a couple of meetings with Ustav informatiky Slovenska akademia vied (UI SAV) it was decided that the Secure Agent Infrastructure (SAI) could help solve this mesh-network situation and ensure that all MBRs have an up-to-date local knowledge.

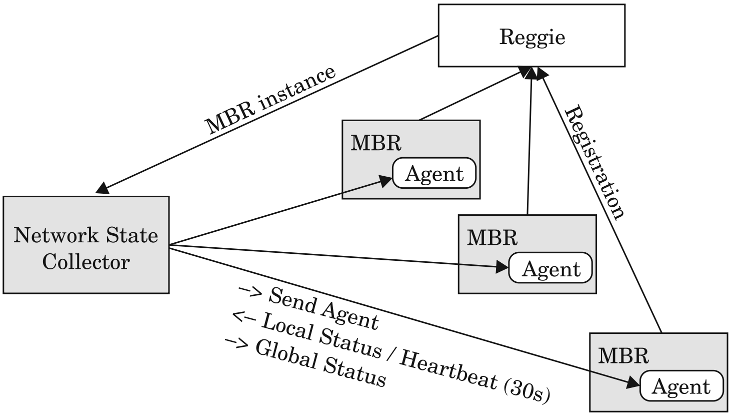


Fig. 1. Overview of SA & MBR Integration

The design was broken into three core sections:

1. MBR Interface – QinetiQ
2. MBR Secure Agent – UI SAV
3. Network State Collector (NSC) – QinetiQ & UI SAV

The MBR Interface would enable communications between the Agent and the MBR to ensure the local knowledge is passed to the NSC and the tailored remote knowledge is passed back into the MBR policy-engine for better decision making.

The MBR Secure Agent would be responsible for going to all registered MBRs on the network and passing the data back and forth between the appropriate locations.

The NSC will take all the information gathered from the MBRs in the network and build a representation of the topology so that it can be tailored to each MBR and returned via the Agent.

Components

MBR Interface

The MBR Interface to the SAI will allow the agent to request what is referred to as MBR local knowledge, i.e. the network state in which that MBR is currently in, the subscribing subnet and local state of bearers. The Agent will ensure each MBR's local knowledge is passed to the NSC to calculate the wider picture of the network. This in turn is used to create the unique remote knowledge for each MBR and the agent will ensure that each MBR receives its personal remote knowledge picture.

For example, consider the simple three node MBR network in Figure 2 below.

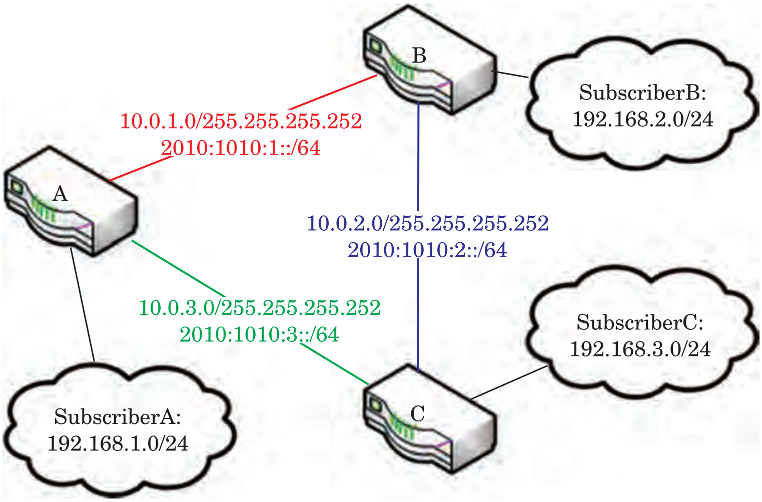


Fig. 2. Three Node MBR Network

MBR A has a direct link with MBR B (red line) and MBR C (green line), but has no knowledge about the status of the link between MBR B and C (blue line). However, if each MBR collates its own knowledge of the network and distributes this to a central server the state of all links can be disseminated to all MBRs.

Given a situation, such that the link between MBR B and C is unavailable and the link between MBR A and B is unavailable.

– MBR A senses its loss of link to MBR B

– MBR C senses its loss of link to MBR B

– MBR B senses its loss of links to MBR A and C

However, MBR A will “expect” that MBR C is still connected with MBR B but has no method of validation, however, with both MBR A and C advertising its local knowledge MBR A can now “learn” that C has not got a connection to MBR B and MBR C would “learn” that MBR A has not got a connection to MBR B. Therefore, neither MBR would allow traffic to traverse the network which is destined for MBR B, thus reducing the amount of bandwidth that is utilised and helping to ensure all accessible services are supportable through the network.

The proposed component would bridge the MBR Orchestration Layer (`orchd`) with the SAI. The orchd application is the core daemon and message bus for MBR operations, enabling components to communicate with each other along with the Linux kernel. The Orchestration Layer utilises a SHA-1 encryption and comprehensive packing functionality to allow secure and compact communications. Therefore to ensure compatibility a new component (`open-interface`) will expose certain MBR functionality to the Agent via specific messages passed over a loopback network socket. This functionality will consist of:

– Register MBR – This function would register the MBR with the “reggie” service (expected to be running with the NSC) to enable Agent communications.

– Retrieve Local Knowledge – This message type returns the specific MBRs current knowledge of the status and network configuration of each of the network interfaces and the serving Local Area Networks. This is one of the two core messages required to enhance the MBRs knowledge, the Agent will collect this information from all MBRs in the network and submit it to the NSC for processing.

– Update Remote Knowledge – This message type accepts the tailored updated remote knowledge calculated by the NSC. This is the second core message type required for the enhancements by giving the MBR an understanding of the topology of the network to know which communication bearers access which served Local Area Networks.

– Change Running Configuration Set – This function would allow the Agent to alter the current running configuration set to another one stored on the MBR. An added bonus being that an Agent could be sent out to all registered MBRs in the network and told to start running a pre-set configuration, such as an emergency configuration to disable any restrictions on network access. Currently only capable through the MBR Web Interface (web-utility) and acts as proof that other web functionality could be ported across.

– Retrieve Policy Table – This function returns a copy of the specific MBRs installed policy table, both IPv4 and IPv6. Another added bonus being that an Agent could be sent out to all registered MBRs and bring back a copy of each policy table to build an exact replica for management purposes. Currently only capable through the MBR Simple Network Management Protocol (SNMP) Interface (network-agent) and acts as proof that other SNMP functionality could also be ported across.

The local knowledge is passed back to the NSC for processing into global knowledge (topology) of the entire network. The NSC then tailors the remote knowledge for each MBR, giving it a personal view of the entire network accessible to them. Taking the previous simple network example (Fig. 2) in a full mesh:

- MBR A can be told:
  - MBR B is reachable via both MBR B and C
  - MBR C is reachable via both MBR B and C
  - MBR B is directly connected via interface one
  - MBR C is directly connected via interface two
- MBR B can be told:
  - MBR A is reachable via both MBR A and C
  - MBR C is reachable via both MBR A and C
  - MBR A is directly connected via interface one
  - MBR C is directly connected via interface two
- MBR C can be told:
  - MBR A is reachable via both MBR A and B
  - MBR B is reachable via both MBR A and B
  - MBR A is directly connected via interface one
  - MBR B is directly connected via interface two

Therefore, if any interface changes state while parts of the mobile mesh network move around the personalised remote knowledge sent to each MBR can be updated. Such as if the link between MBR B and C is unavailable, due to if either or both MBRs disconnected from a shared/multi-nodal network such as Wi-Fi or Internet, the tailored remote knowledge for each MBR would change to become:

- MBR A would be told:
  - MBR B is reachable via only MBR B
  - MBR C is reachable via only MBR C
  - MBR B is directly connected via interface one
  - MBR C is directly connected via interface two
- MBR B would be told:
  - MBR A is reachable via only MBR A
  - MBR C is reachable via only MBR A
  - MBR A is directly connected via interface one
- MBR C would be told:
  - MBR A is reachable via only MBR A
  - MBR B is reachable via only MBR A
  - MBR A is directly connected via interface one

If and when MBR network interfaces change state they immediately inform the NSC via the Agent a quicker response to network changes should be achieved, maybe not in the region the MBR is used to (< 20 ms) but considerably better than the dynamic routing protocols available today (~ 40s) and without over flooding the network with adjacency messages along with other issues that routing protocols face in an ever changing environment (e.g. designated and backup router assignment).

Even with this regularly updated knowledge ultimately a timeout would still have to exist for assistance in capturing against ambiguous groups forming. For example taking the original setup where MBR B is completely disconnected from MBR A and C, as long as MBR A and/or C has connectivity to the wider mesh network, the NSC will process the updates that MBR A and C send when they locally sense the loss of those interfaces connection to MBR B. However, if MBR A and C have formed a separate group without access to the NSC then the original problem would still exist; this is due to the centralisation/server-style of the NSC.

Secure Agents

The MBR Agent has been developed using the framework and standards common for the whole SAI. The standard envelope of a secure agent has been used, and filled with code specific to the task of communicating with a MBR via its socket-based interface.

Figure 3 shows the class diagram of the final MBR Agent implementation. The class `MbrAgent` is derived (as said above) from the standard SAI Agent envelope (GATIAL et al. 2011). It uses a structure of several support classes, which are abstracted into several interfaces – most notably `MBRStatusInter-`

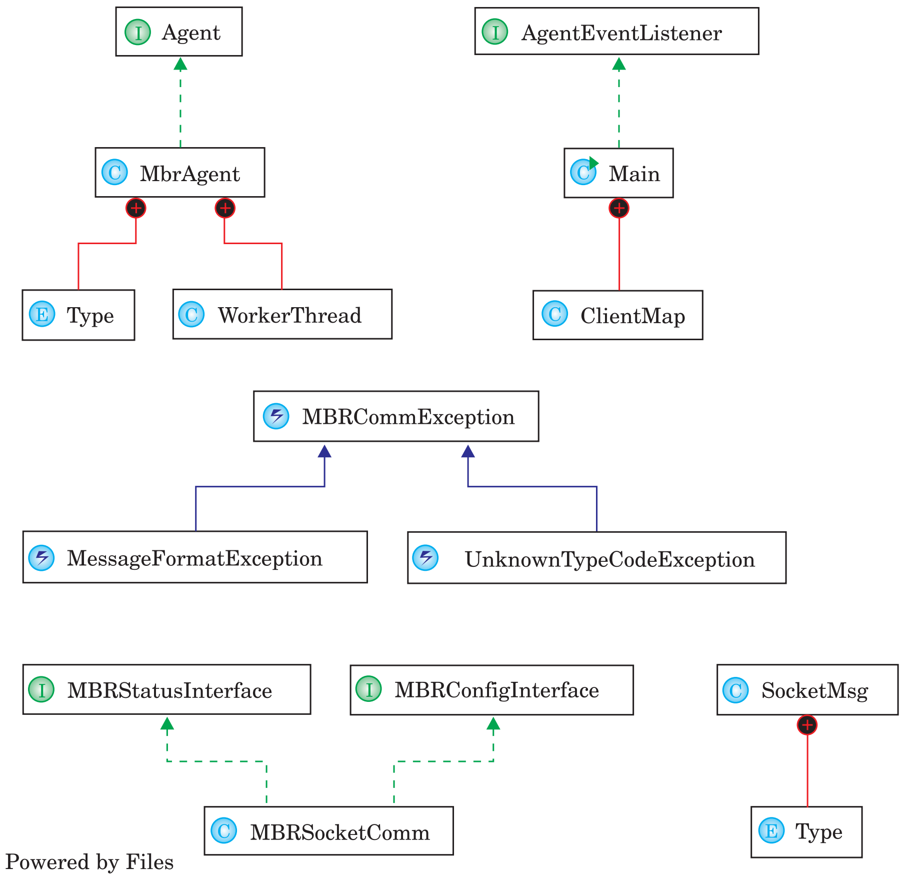


Fig. 3. The class diagram of MBR Agent implementation

`face` (methods for querying MBR status) and `MBRConfigInterface` (methods for changing MBR configuration), which are both implemented in the MBRSocketComm class, which gives these methods actual communication transport – via sockets to the MBR Interface. Exception handling is done via a small hierarchy of `MBRCommException` and its implementing classes.

Testing of a SAI agent is a complicated task, since agents are code objects which do not act on their own, but upon request. To allow for proper testing of the code which implements the agent, a separate envelope in the form of a common Java program has been developed (`eu.secricom.dsap.agent.mbr.test.MbrTests`) and this has been linked to the same classes, which implement the actual `MbrAgent` class. This program can be run using command-line arguments, and so the tests may be automated. Other methods of automated testing have not been used, since they are not standardised in the SAI development environment, and would create too much complexity.

A placeholder implementation of the MBR interface has also been created, in order to be able to autonomously test the MBR Agent communication. This may be found in the class `eu.secricom.dsap.agent.mbr.test.DummyServer`.

Only after the tests of the MBR Agent have been done using this artificial testing grounds, a real `MbrAgent` implementation has been successfully tested with the real MBR Interface.

Network State Collector

The learning process is provided by the NSC component itself which gathers together the different fragments of local knowledge coming from the MBRs and analyses them to work out the inherent connectivity. By performing an iterative search process that "joins up" the details about active interfaces, the NSC is able to determine, for each MBR, the possible networks accessible from any one interface. Figure 4 illustrates the effects that network topology can have on the "reach" of an interface, i.e. what other network nodes can be accessed from a particular interface.

Network A in the example is a fully "looped" arrangement of connections that allows any interface on any node to route its traffic to all other nodes. Network B illustrates that interfaces marked as C1 and C2 may only broadcast to nodes A and B respectively; likewise for interfaces D1 and D2 on node D which can only reach nodes E and F.

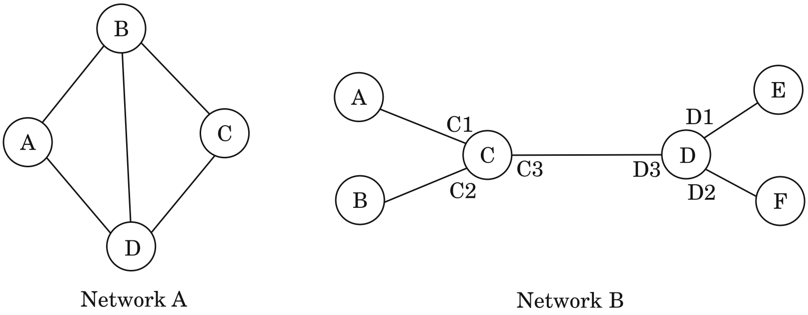
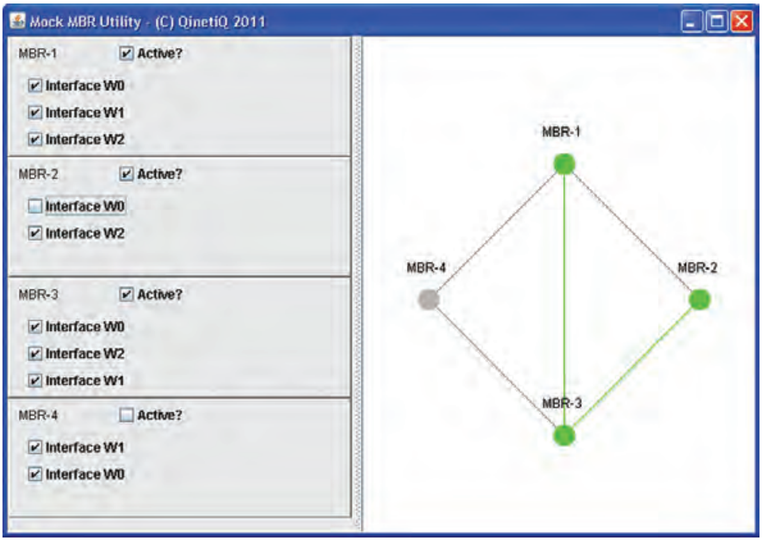


Fig. 4. Examples of interface "reach"

The NSC needs to inspect the network as a whole to draw up a picture of the possible routing choices that will be used to build the global knowledge passed to still-alive and reachable MBR's.

The development of the Network Monitoring system for the MBR was implemented in parallel. QinetiQ took the lead on developing the MBR Interface along with the NSC. UI SAV took the lead on developing the SA for this purpose. Further use-cases were discussed around the MBR Interface component, such as the ability to get and set certain MBR properties.

The development of the NSC took place in two phases. During the initial phase, a bespoke tool (see Figure 5) was created to stand in place of the "live" MBR agent, which was still being developed in parallel. This tool was capable of being configured to show a variety of network forms and also the connectivity required between the MBRs.



Mock MBR Utility - (C) QinetiQ 2011

MBR-1 ☑ Active?
☑ Interface W0
☑ Interface W1
☑ Interface W2

MBR-2 ☑ Active?
☐ Interface W0
☑ Interface W2

MBR-3 ☑ Active?
☑ Interface W0
☑ Interface W2
☑ Interface W1

MBR-4 ☐ Active?
☑ Interface W1
☑ Interface W0

Fig. 5. Bespoke "mock" MBR tool

Communication between the mock MBR tool and the NSC was achieved using network socket architecture. The tools interface allowed for the "state" of an illustrated MBR to be altered by turning on and off different aspects of its functionality. Whenever this would happen, a new set of local knowledge was sent to the NSC which would re-assess the situation and define a fresh set of global knowledge data. This ability to precisely choose the network state meant that it was possible to very easily validate the functionality of the NSC component.

It was decided to replicate the network graph drawing within the NSC also in the first phase of development. This provided the enormous benefit of being able to perform a comparison between the diagrams shown in the two different tools. As the network state in the mock MBR window is modified, the NSC was shown to "catch up" with its own view of the network;s revised infrastructure based on a change in broadcast local knowledge.

The NSC can use its complete knowledge of the network to build tailored remote knowledge unique to each MBR registered in the network. Ideally it will do this quickly enough that the instance an updated local knowledge message comes from an MBR the NSC will be able to recalculate the topology of the network and send updated tailored remote knowledge back to each MBR.

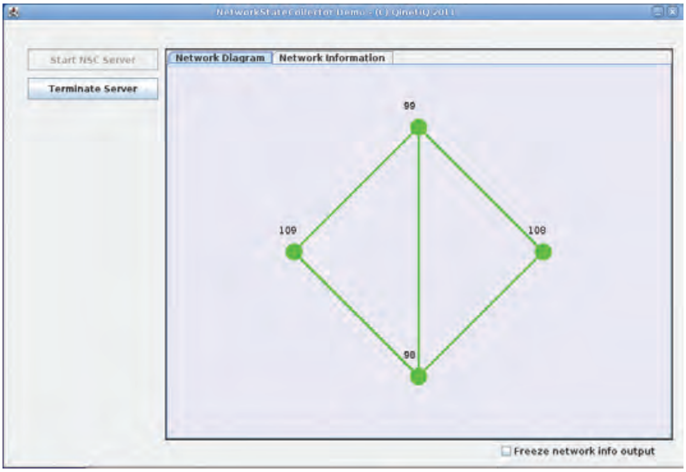


Fig. 6. Network state collector application

Having satisfactorily completed the logic and implementation of the NSC, the second phase of its development involved replacing the socket communication mechanism with the SAI toolkit functionality.

Currently, the NSC is designed to work with a "mock" or "fake" MBR service that can supply details about MBR devices on the network. This is handled by a `ServerProcess` class and it is this class which would need to undergo most of the changes to make use of the SAI toolkit as required.

First of all however, the actual agent "definition" would need to be added into the NSC project. An agent item is a component that can be dispatched to a location (an MBR in this case) that makes use of the Distributed Secure Agent Platform (DSAP) service. This service would receive the agent and allow it to execute in its "address" space. For this work package, UI SAV are creating a "layer" that would exist on an MBR and would act as the go-between for the MBR and the NSC. Note that the agent definition (also being created by the UI SAV) is based on a subclass of Agent which is part of the toolkit and would enter into a permanent loop (via the toolkit `AgentThread` class) polling the MBR device to acquire local knowledge.

This agent component has to be deployed initially by the NSC itself when the NSC first initialises and detects the presence of running MBRs. This would be done by adding a small element of start-up code to the NSC which would look for the "reggie" service that should already be active on the network. This service can be used to obtain the details about the registered MBRs (when an MBR device starts up, it would need to locate the "reggie" service and register its existence).

For each "discovered" MBR, the lookup-client `upload` function would need to be called to send an instance of the agent to the MBR where it would start executing. The agent would presumably call its `fireEvent` method to send data to the NSC about local knowledge and heart-beat messages.

Within the `ServerProcess` of the NSC, the `notify` method is used to receive the messages coming in from the deployed agents. Once the data from an agent is received, it is handled in exactly the same way as now. If global knowledge needs to be created by the NSC, it can be passed back to the agent using the appropriate SAI toolkit message.

Conclusion

One immediate issue that sprang from the use of the "mock" MBR system to mimic the required behaviour and infrastructure was the impact of the routing protocol. During tests of the system working with real MBR's, when an MBR link was disabled, it broke the communication with the MockMBR component (and ultimately with the NSC). Once the routing protocol had resolved an alternative route, the NSC was able to once again determine the change in local knowledge and so compute new global knowledge. However, it takes quite some time for this process to be performed so that the NSC can resume its full function. It is hoped therefore to demonstrate with the use of the SAI toolkit, this problem is alleviated. It would be interesting to see if this large amount of latency is eliminated by using another layer of communication which is presumably fundamental to the toolkit and sits below the routing protocol.

Functionality that exposes the information for remote processing helps further on the reduction of hardware requirements for the MBR enabling the software to easily operate on hardware constraint mobile devices.

The ability to combine the exposing functionality from the Web and SNMP Interfaces in a unified secure manner is a new opportunity for remote management and control of many MBR aspects. Once full integration with the

SAI has been completed it will be a simple case of adding new message type functionality to the open-interface MBR component and creating an Agent with a new `fireEvent` method.

Recommendations

The core recommendation would be to de-centralise the NSC, so that each MBR can initialise its own operating knowledge directly with other MBRs and not reliant on a single NSC. This should also help in keeping seamlessness and reducing delay in passing tailored local knowledge out to all relevant MBRs.

Further to de-centralisation of the server, by combining the MBR knowledge and routing information dissemination the system could reduce its network bandwidth footprint and ensure that no information is unnecessarily duplicated. To make this easier the QinetiQ Routing Bridge (SPENCER et al. 2009) (a patented proprietary routing protocol) could be utilised in assisting the translation between routes and remote knowledge.

Translated by AUTHORS

Accepted for print 30.06.2012

References

GATIAL E., BALOGH Z., ŠIMO B., HLUCHÝ L. 2011. *Distributed secure agent platform for crisis management*. In: SAMI 2011: 9th IEEE International Symposium on Applied Machine Intelligence and Informatics. Budapest, IEEE, p. 247-253. ISBN 978-1-4244-7428-8.

SPENCER J., KENDRICK G., MILLINGTON P., STEPHENSON I., ENTWISLE P. 2009. *Communications System QinetiQ Limited*, December 2008, WO 2009/001041 A1.

Reviewers

Zoltan Bologh, Jarosław Bosy, Apostolos P. Fournaris,
Jacques Fournier, Emil Gatial, Gabriela Gheorghe, Bernard Kontny,
Marian Kopczewski, Foued Melakessou, Jan Pawlak, Adam Rybka,
Jaroslaw Teska, Gyula Tóth