Novel DSA Algorithms and Virtual Machines for Software Defined Radio

ALF BENGTSSON, PATRIK ELIARDSSON, HUGO TULLBERG,
PETER STENUMGAARD, KIA WIKLUNDH

FOI, Swedish Defence Research Agency, is a mainly assignment-funded agency under the Ministry of Defence. The core activities are research, method and technology development, as well as studies conducted in the interests of Swedish defence and the safety and security of society. The organisation employs approximately 1000 personnel of whom about 800 are scientists. This makes FOI Sweden's largest research institute. FOI gives its customers access to leading-edge expertise in a large number of fields such as security policy studies, defence and security related analyses, the assessment of various types of threat, systems for control and management of crises, protection against and management of hazardous substances, IT security and the potential offered by new sensors.

FOI
Swedish Defence Research Agency
Information Systems
Box 1165
SE-581 11 Linköping

Phone: +46 13 37 80 00
Fax: +46 13 37 81 00

www.foi.se

FOI-R--2919--SE
ISSN 1650-1942

Scientific report
November 2009

Information Systems

Alf Bengtsson, Patrik Eliardsson, Hugo Tullberg,
Peter Stenumgaard, Kia Wiklundh,

Novel DSA Algorithms and Virtual Machines for Software Defined Radio

Titel	Nya DSA-algoritmer och virtuella maskiner för mjukvarudefinierad radio
Title	Novel DSA Algorithms and Virtual Machines for Software Defined Radio
Rapportnr/Report no	FOI-R--2919--SE
Rapporttyp Report Type	Vetenskaplig rapport Scientific report
Månad/Month	November
Utgivningsår/Year	2009
Antal sidor/Pages	52 p
ISSN	ISSN 1650-1942
Kund/Customer	Försvarsmakten
Kompetenskloss	22 Robust Telekommunikation
Extra kompetenskloss	26 IT-säkerhet
Projektnr/Project no	E7138
Godkänd av/Approved by	Jacob Löfvenberg Magnus Jändel

FOI, Totalförsvarets Forskningsinstitut
Avdelningen för Informationssystem
Box 1165
581 11 Linköping

FOI, Swedish Defence Research Agency
Information Systems
Box 1165
SE-581 11 Linköping

Sammanfattning

Det är troligt att i en nära framtid kommer frekvensreglering att genomgå en förändring på så sätt att vissa frekvensband övergår från att vara statiskt planerade till att vara dynamiskt tillgängliga. Dessa band kommer vara föremål för så kallad dynamisk spektrumaccess (DSA) vilket innebär att de delar av spektrum som inte används för tillfället kan användas av andra användare. Huvudfokus för projektet Dynamiska Telekommunikationslösningar är att undersöka Försvarsmaktens möjligheter att dra nytta av DSA-utvecklingen i framtida radiosystem.

Projektet bedriver forskning om metoder för kognitiv radio som känner av signalmiljön i frekvensband och använder denna information för att justera radioparametrar dynamiskt. Avkänning av signalmiljön kan göras autonomt för en enskild radionod eller som samverkansuppgift där flera noder bistår varandra genom att utväxla data. Oavsett strategi i detta avseende kan avkännandet av miljön bli påverkad av en fientlig insats. Av det skälet är det mycket viktigt att metoder för kognitiv radio i detta avseende tar hänsyn till säkerhetsaspekter för att undvika sådana attacker. I denna rapport presenteras resultat från arbetet med säkerhetslösningar och med algoritmer för dynamisk frekvensallokering.

Förslag på säkerhetslösning presenteras. Vi föreslår en arkitektur baserad på virtuella maskiner. Den centrala egenskapen är förmågan hos en virtuell maskinmonitor att separera och isolera delar inom systemet. Detta kan användas för att separera olika klasser av information. En ny algoritm för dynamisk frekvensallokering har utvecklats i projektet. I rapporten visas hur denna algoritm får bättre prestanda än traditionell spektrumallokering som baseras enbart på mätning av störsignalers effekt i ett aktuellt frekvensband. Denna algoritm och eventuellt flera kommer att implementeras i en demonstrator för att utvärdera prestanda i en verklig radiomiljö.

Nyckelord: dynamisk spektrumaccess, mjukvarudefinierad radio,
frekvensplanering, virtuella maskiner, hypervisor, Xen, separation kernel,
mandatory access control,

Summary

It is foreseen that in the near future the regulation of radio spectrum will undergo a change from a static and slow process to a more dynamic and faster process know as Dynamic Spectrum Access (DSA). In a DSA environment parts of the available spectrum will be made available for dynamic use of instantaneously unutilized frequencies. The main focus of the project Dynamic Telecommunications is research concerning methods for a cognitive radio (CR) to sense the environment, e.g. utilization of frequencies, and to use this information to adjust parameters of the radio. The sensing of the environment might be done autonomously by a radio node, or it might be a collaborative task of many radio nodes, helping each other by exchanging data. In either case, there are security issues, since the sensing could possibly be disturbed by a hostile opponent. Therefore, it is important to have security measures, intrinsic in the cognitive radio node, to mitigate such attacks.

In this report, results are presented from the work on security measures and algorithms for dynamic frequency allocation.

We propose an architecture based on Virtual Machines. The central security quality is the ability of a virtual machine monitor to separate and isolate partitions within the system. This can be used to separate different classes of information.

Novel DSA algorithms have been developed in the project. These algorithms are shown to outperform traditional algorithms based on simple interference energy detection only. The algorithms will be implemented in a demonstrator to evaluate the performance over a real radio channel.

Keywords: dynamic spectrum access, software defined radio, frequency planning, virtual machines, hypervisor, Xen, separation kernel, mandatory access control,

Table of Contents

Part 1

1 Introduction

Emerging broadband wireless services demand access to increasing amounts of radio spectrum. International Mobile Telephony Advanced (IMT-Advanced) envisions peak data rates up to 100 Mbps for high mobility, and 1 Gbps for low mobility, using up to 100 MHz bandwidth [1].

Current spectrum allocation does not allow such bandwidths to be allocated, and thus the radio spectrum is often perceived as a scarce resource. However, it has been observed that spectrum assigned to licensees is not fully utilized [2].

Traditionally, frequency planning is done at a centralized level by the regulation authorities at international and national levels. Internationally, spectrum allocation is coordinated by ITU-R's World Radiocommunication Conference (WRC) which is held every two to four years. This regulatory process is often criticized for being slow and static, which delays the development and introduction of new communication systems.

However, over the last decade the markets for electronic communication have been opened up to competition and the relation between regulators, operators and developers of equipment is no longer as close as it has been. The technical development is generally heading in the direction of smarter and more adaptable systems and solutions [3][4].

It is foreseen that in the near future the regulation of radio spectrum will undergo a change from a static and slow process to a more dynamic and faster process know as Dynamic Spectrum Access (DSA). In a DSA environment parts of the available spectrum will be made available for dynamic use of instantaneously unutilized frequencies.

A Cognitive Radio (CR) system [5][6] is a radio and related infrastructure that is able

(a) to detect users' communication needs as a function of use context, and

(b) to provide the radio resources and wireless services most appropriate to those needs.

A CR should hence be capable of machine learning, and Mitola considers this as an essential and integrated part of CR [5]. SDR Forum on the other hand does not include machine learning in its definition of CR [7], and essentially considers a CR to be a radio capable of DSA. Examples of initiatives where SDR and CR are used as enablers for DSA are Wireless Access Policy for Electronic Communications Services (WAPECS) [17] and White Space Coalition [18].

Even in a DSA environment the use of spectrum will be more or less regulated to limit the interference between DSA-capable radio and to non-DSA-capable systems. Such regulations can be rules, policies and voluntary agreements.

However, it should be noted that the level of interference can be significantly higher in a DSA environment, since radios will access radio spectrum in a non-coordinated way. Thus, DSA-capable radios must be able to handle this increased interference level.

It is envisioned that a cognitive radio system, which autonomously senses the electromagnetic signal environment and is aware of the user's needs, can perform a negotiation with the spectrum owner regarding access to and use of spectrum and thereby provide a more efficient utilization of the radio spectrum.

Current research focuses on detection methods, e.g. [8], and not so much on spectrum measurement techniques. In DSA environments new methods for spectrum measurements and policies are needed for several reasons, e.g.:

- DSA applications require new methods that are "ambient-noise limited" in which the total signal environment is considered as compared to current "interference-limited" methods where only other users are considered in the intersystem-interference analysis.

- The intersystem-interference analyses cannot be performed in advance for a limited number of static cases as is the case today. In a DSA environment the number of possible intersystem-interference cases will be too large.

- Real-time analysis requires analysis models with only a limited number of signal and interference parameters and thus a simple measure of the total instantaneous interference is needed.

A common measure of the instantaneous signal and interference level is the so called "interference temperature" which has been proposed by the Federal Communications Commission (FCC) [9]. The interference temperature is simply a measurement of the total RF power generated by undesired emitters plus noise sources that are present in a receiver system per unit bandwidth. However, it is well known in the interference research that one difficulty with such an approach is that the wave form, not only the power, of an interfering signal can significantly affect the performance of a disturbed system. Thus, the interference temperature metric could be too blunt and must be further investigated to determine the risks of under/overestimation of the interference impact if used.

In this work we develop new spectrum sensing and classification algorithms. We use the Amplitude Probability Distribution (APD) concept [13] and signal models with reduced parameter sets to classify the interference environment and perform parameter estimation for non-Gaussian, impulsive channels.

If the impulsiveness is correctly identified, then the channel can be efficiently used (in some cases even better than under a Gaussian assumption, cf. the 2-state Gilbert-Elliott channel model or generalizations to n states [14][15][16]).

When the instantaneous channel is identified, the controlling entity, “cognitive engine”, can decide on the appropriate waveform to use e.g., OFDM with Adaptive Coding and Modulation (ACM) for dynamic bit allocation to match instantaneous transmission rate to capacity. This property makes OFDM an attractive waveform and it is used in LTE and LTE-Advanced [11], IEEE 802.22 [12], as well as in the ubiquitous WLAN standard IEEE 802.11x.

2 System Model

In this section we briefly describe the scenario and system assumptions used in this work.

2.1 Scenario

The purpose of the scenario is to provide a means for algorithm development, simulation and demonstration. The scenario should be realistic (complex) enough to capture real opportunities and challenges, yet simple enough to provide understanding and ease of use. The scenario is intended to be used first in baseband (or passband) simulations and then migrate to hardware demonstrations. It will be expanded to include more elements as we gain experience. The scenario is depicted in Figure 1.

The basic scenario consists of:

- 2 independent point-to-point (P2P) communications links[1]
- 1 Primary source e.g., TV transmitter[2]
- External interference source(s) that may be partial band, partial time, and time-varying
- Internal interference, i.e., effects of own spectrum use

We may make a distinction between Opportunistic Spectrum Access (OSA) and Dynamic Spectrum Access (DSA). In the case of OSA a system utilizes, in a secondary manner, spectrum which is licensed to another primary user. I.e., the system must vacate the spectrum when the primary user becomes active. Examples of primary systems are TV transmitters, radars, and cell phone systems. In DSA there is no primary user and all users access spectrum on an equal footing, and spectrum sharing should be governed by policy or negotiations. In the case of OSA, policy application or negotiations can be necessary between multiple secondary users.

The primary user in the model provides opportunity for OSA, i,e., we can access spectrum not licensed to us to momentarily increase the data-rate. The primary user can be a TV transmitter or a radar. A TV transmitter is a “high power, high tower” user and provides approximately the same spectrum usage over a large geographic area. The TV transmission should be fairly easy to detect and/or predict (stationary, often known, location, no covert operation). Causing

[1] Future extensions include more, possibly inter-dependent P2P links, multi- or broadcast.
[2] Future extensions include additional sources such as radar tranceivers.

interference to the TV transmitter cause nuisance to the viewer but is not life-threatening. Radars may be lower power, lower tower, and thus harder to locate. The transmission pattern of a mechanically sweeping (rotating) radar should be easy to predict. We may however have to consider the back lobe as well as the main lobe in order to avoid interfering with the radar. Interfering with the radar may have grave consequences.

Additional elements in the scenario can be other secondary users. The presence of such requires us to consider policy application or negotiation. In a first version of our scenario we do not consider other secondary systems.

External interferences emanates from industries, consumer electronics such as computers, microwave ovens, and other sources. These interferences may be full or partial band interference, constant or intermittent, and time-varying.

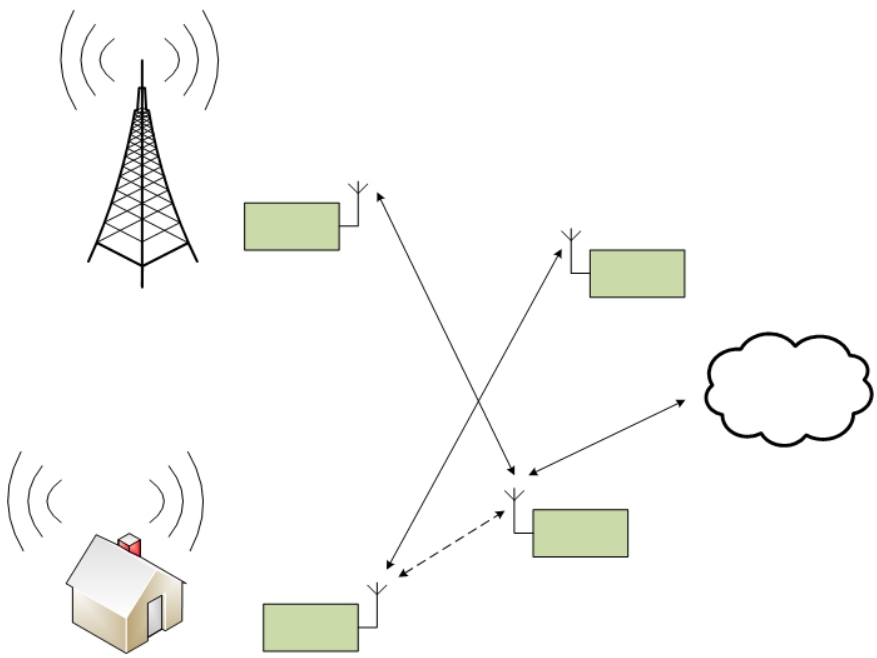

Figure 1 The two (three) P2P links experience interference from a primary user (TV), an external interference source (factory), and from internal sources (other terminals in the system). The cloud is intended to illustrate that all the terminals can be parts in a larger radio network.

2.2 System Assumptions

The communication system is using Orthogonal Frequency Division Multiplexing (OFDM). OFDM allows for adaptive coding and modulation and can hence be adapted to varying channel conditions. The parameters of the OFDM system, such as the time-frequency grid, the length of the cyclic prefix etc. depends on the application at hand.

OFDM systems are also used in civilian high data-rate systems, such as 3GPP Long Term Evolution (LTE) [11] and IEEE 802.22 [12].

LTE can handle bandwidths up to 20 MHz. In LTE-Advanced carrier and spectrum aggregation is used to provide bandwidths up to 100 MHz over possibly non-contiguous bands. In the case of carrier aggregation, multiple adjacent component carriers are collected into an overall wider bandwidth. E.g., five 20 MHz component carriers give a total bandwidth of 100 MHz. In spectrum aggregation the carriers need not to be adjacent [11].

IEEE 802.22 is a standard for Wireless Regional Area Network (WRAN) utilizing white spaces in the TV frequency spectrum on a non-interference basis [12]. The intent of this standard is to provide broadband access in rural and remote areas with performance comparable to DSL and cable modems. Cognitive Radios (CR) will reuse fallow TV spectrum in an opportunistic way by detecting if the channel is occupied before using it.

2.3 Protocol Issues

There are several issues related to signalling and communication protocols that must be determined in a real system. Examples include

- Access method (CDMA, FDMA, TDMA, CSMA, etc.)
- Whether there is a centralized control or not
- Whether the nodes signal between each other or if the system relies on adherence to policies
- The real-time requirements, the trade-off between fidelity and time (delay).

These issues will be addressed in the implementation of the demonstration system.

3 Cognition Node

In various communication systems it is possible to change for example transmitter power, frequency of operation, modulation scheme, battery usage and processor usage to maintain a communication link between two or more nodes. However, these parameters tend to change in a planned way, e.g., the modulation scheme may switch to a lower throughput if the quality of the link drops under a certain threshold to maintain the link. With a cognition node[3] in the radio these changes can take place in an unplanned fashion instead. To achieve this unplanned behaviour, the cognition node has to be aware of its environment, its own capabilities, and have the ability to learn what to do and judge what is advantageous and disadvantageous.

3.1 Environment Awareness

The environment that the cognition node has to be aware of can be further divided into three separate fields; namely the radio domain, the policy domain and the user domain [29][30]. These three domains affect each other and the cognitive radio's behaviour as shown in Figure 2.

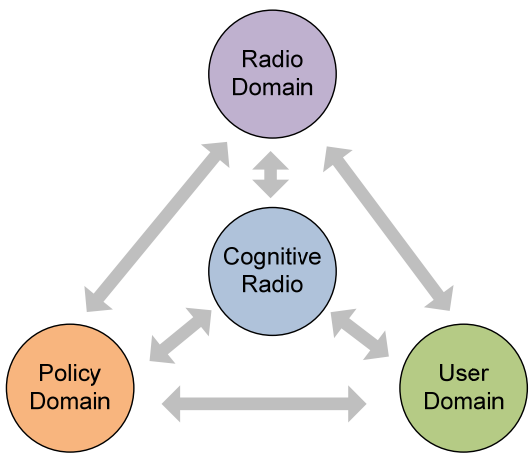

Figure 2: The three different domains of which the cognition node has to be aware.

[3] Some authors refer to the cognition node as a cognitive engine. We consider the cognitive engine to be a function performing reasoning and learning, whereas the cognition node also considers policies and their description.

The user domain defines the requirement that the user's application has on the radio; some services can have very high priority to reach the destination immediately whereas other services require a very low Bit Error Rate (BER) while the transfer time is not so important etc. This can be summarised as user quality of service (QoS).

The radio platform and the radio environment are two different fields that are included in the radio domain. The radio environment is more or less everything that is outside the radio, like spectral activities from other systems. The radio platform describes the functionality of the radio, battery-driven, available waveforms, adjustable parameters in the waveforms, available antennas etc.

In the policy domain radio regulations are modelled to ensure that the radio follows interference levels, does not use radar spectrum for communication (because the radar spectrum seems to be unused but the radar is actually waiting for echo) etc. Policies can be a set of rules to follow. If the radio encounters a specific state a specific set of rules have to be followed. The policies do not specify how the rules should be followed, that is the cognition node's task to solve.

For DSA applications, spectrum opportunity detection algorithms in the radio domain are crucial, but the definitions of policies in the policy domain are at least as crucial. Without well defined policies the radio could start to use forbidden frequencies and interfere some system which could lead to disasters, the radio infrastructure for the air traffic control and radar systems for example.

In a cognitive radio perspective it is of interest to classify the detected primary users. The goal of the classification is to detect what kind of waveforms the primary users have, so that the cognitive radio can establish a connection to them.

3.2 Reasoning and Learning

When the cognition node is aware of its surrounding environment and its own capabilities, it can start to learn and change the radio parameters in an unplanned fashion. The algorithms used for the implementation of the reasoning and learning mechanism in the cognition node can be inspired from several areas [29][31][32]. Some of them are:

- Decision Tree
- Case Based Reasoning
- Genetic Algorithms
- Game Theory
- Iterative Waterfilling

To describe all these algorithms in detail is beyond the scope of this report.

A decision tree is a basic way of determining the radio parameters. The decision tree is a predetermined set of choices and is constructed when the radio is designed. Although the tree may be complex, it will not be autonomously modified by the radio itself.

In a case based reasoning algorithm the current state of the radio is described with a problem function. A similarity function describes how well the past problem functions match with the current problem function and the utility function calculates the utility metric from the feedback of the decision. A case based reasoner utilizes all these three functions/metrics to come up with a new solution for the current problem function.

Genetic algorithms are based on an evolution process similar to which exists in a living organism. The evolution is typically an iterative process where solution proposals mutate, inherits and the most fitting are selected and survive. A chromosome (problem function in the case based reasoning algorithm) in the cognition node describes the current state of the radio, the current parameters from the different domain that the cognition node is aware of. A set of chromosomes creates a population from which a fitness function selects a subset of chromosomes to form a fitness space. Then the chromosomes in the fitness space mutate and inherit until they have formed a new population (solution).

With the game theory approach each radio in a network is modelled as rational player that wants to maximize its own performance. To be able to share the resource among the players, each player has to take into consideration what the other players' actions are and how it should act to reach the steady-state that is most advantageous for all players.

In iterative water-filling the available frequency spectrum is divided into sub-bands, and a function is used to describe the assigned power level of each sub-band. If the power level is lower then a threshold value, the transmitter adds power into that sub-band until it reaches the maximum power level. The sub-band is considered as bad if the power level is above the same threshold and then the transmitter leaves this sub-band untouched.

Iterative water-filling differs from the other algorithms in the sense that it is not learning from previous decisions and it is only suitable for DSA while the other algorithms can be extended to utilize more information than spectrum holes.

3.2.1 OODA loop

The Observe-Orient-Decide-Act (OODA) loop was originally created and used as a decision model for fighter pilots. It has later been adapted to several other applications in business and learning processes and can also be applied to the decision process of a cognition node. For instance, the spectrum sensing is typically the *observe process* in the OODA loop. The information gained in this process is analyzed (the *orient process*). Based on this information, decisions can be made (*decide process*) and actions can be taken in a certain way (*act process*). An example of an OODA loop for the cognition node is given in Figure 3. Note that in the figure we explicitly show the *learning* and *planning processes*. This may be assumed to be included in the *orient* and *decide processes* of the original OODA loop.

Note that the OODA loop is a continuously ongoing process. Decisions and actions will react in effects on the environment and the environment (including other radios) will react to our actions. We will continuously make new observations, analyses, decisions and actions. The decision process must be able to handle incomplete information. In the case of military communications, the OODA loop must be able to identify and handle electromagnetic jamming and other forms of misleading information.

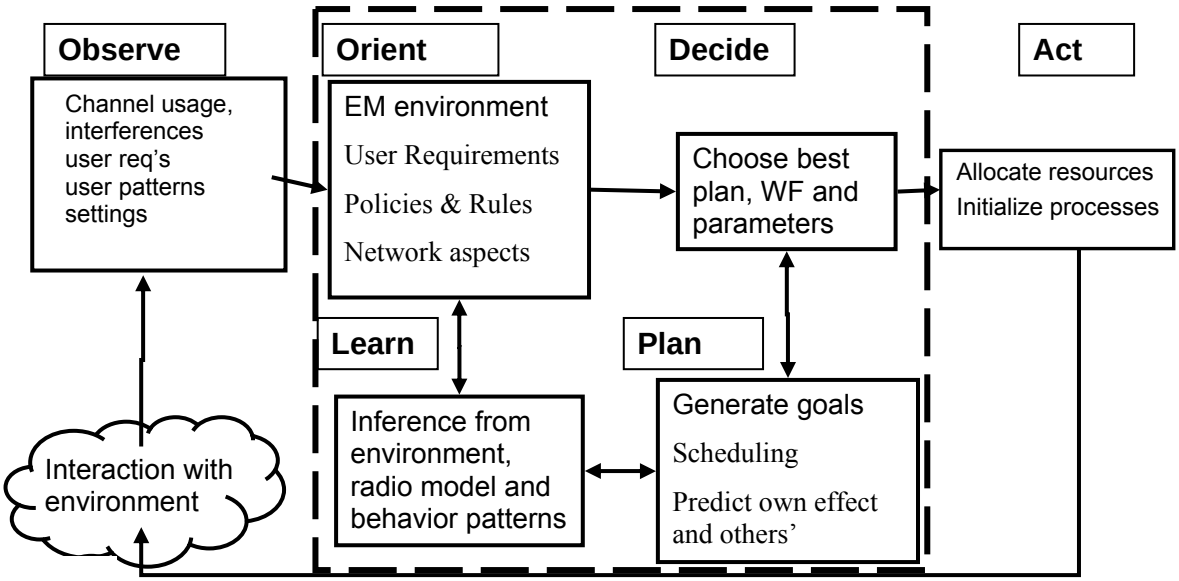

Figure 3: Example of an OODA loop for a cognitive radio.

3.2.2 Frequency Allocation Using a Decision Tree

In the following we will look at an example of a simple decision tree. We will consider a (secondary) user that wants to use a certain frequency band and we study how the decision process can look like. The decision tree is illustrated in Figure 4.

Before usage of the spectrum, the user investigates whether there are any (primary user) existing radio access technologies (RATs) within the spectrum. If the spectrum is used by some RATs, then the secondary user can use the frequency band if it can assure that it will not degrade the quality for the existing RATs.

If there are no RATs, the user will begin to investigate whether this frequency band will give an acceptable quality or not for the intended communication service. The user will first perform an energy detection of the frequency band. This investigation gives a course analysis regarding whether the frequency band is too crowded or not. If this analysis states that there is frequency space available, the user can begin to use the band. If the initial energy detection

indicates that the frequency band might be occupied, a further analysis is performed to determine the occupation more in detail. If it from this further investigation turns out that there is space available, the user can begin to use the band. If there is not enough space for the intended communication, it is necessary to decide if the communication should be performed or not. If the communication is crucial then the decision can be to perform the communication, no matter of the quality. For situations where the communication is not crucial, the user should wait until the frequency band is less occupied.

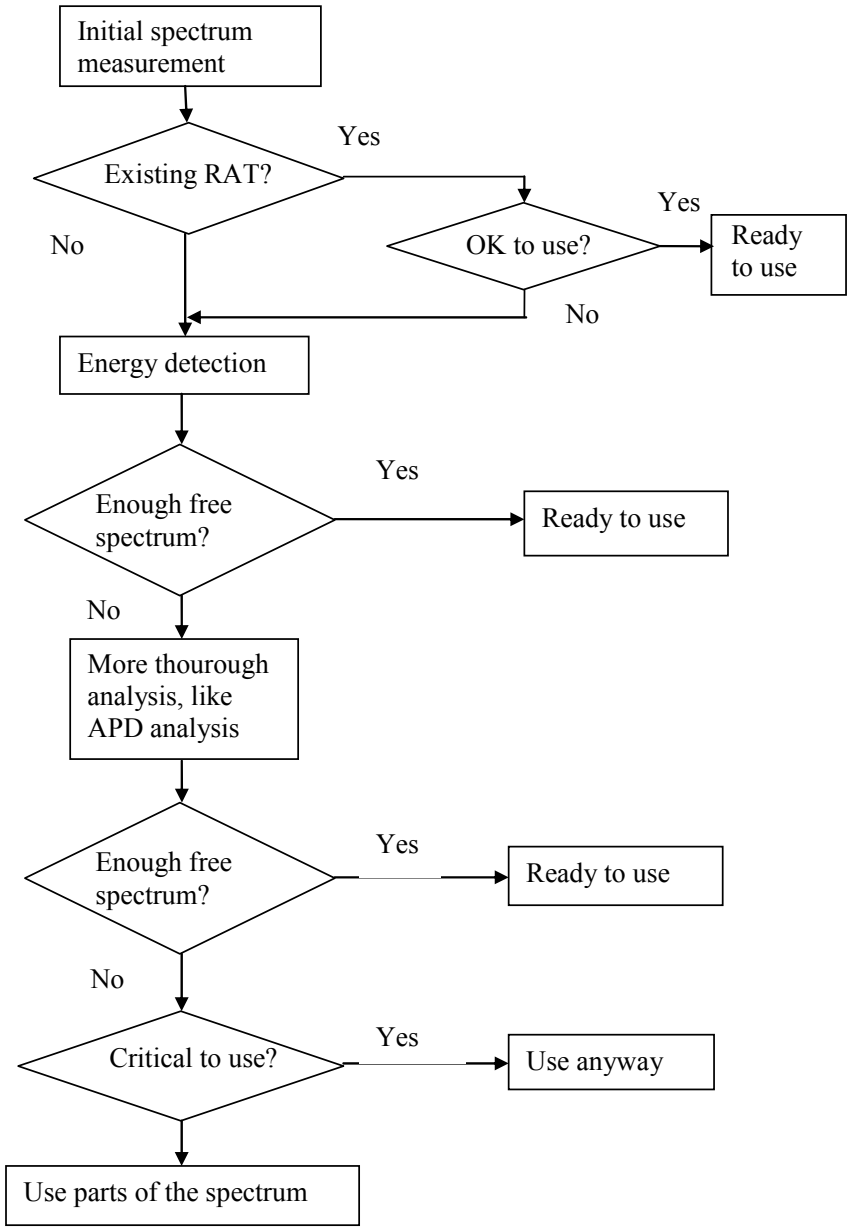

Figure 4: Example of considerations taken for usage of spectrum expressed as a decision tree.

4 Channel Models for Cognitive Radio

The communication channel of a Cognitive Radio is time varying due to the variable nature of the signal environment. The channel can be modelled in several ways. One of the simplest is a Binary Erasure Channel (BEC) with variable erasure probability. Other models include partial-band jamming models [34], Block-Fading Channel models [35] and Finite-State channel models [14][15][16]. A common interference model in the electromagnetic interference literature is the Middleton Class-A interference [19]. This model has the advantage that it can represent a number of interference signals with arbitrary impulsiveness content. By choice of model parameters, we can model a large class of interference ranging from purely Gaussian distributed noise to highly impulsive interference. A communication channel experiencing Class-A interference is referred to as an Additive White Class-A Noise (AWCN) channel [20]. The AWCN is a finite-state model with a particular state description.

4.1 Channel Capacity of the AWCN Channel

To determine the possible performance of a CR operating over an AWCN (the channel model chosen for this work), we give an expression for the channel capacity of the AWCN (the full derivation is found in [36]). If both the transmitter and receiver have knowledge of the channel state, it is straightforward to compute the channel capacity for a finite state channel. Let C_m denote the capacity of the AWGN channel in state *m*. By a time-sharing argument, we get the average channel capacity *C* as [22]

$$C = \sum_{m=0}^{\infty} \pi_m C_m$$

where π_m is the steady-state probability distribution of the Markov chain,

$$\pi_m = e^{-A} \frac{A^m}{m!}, \quad 0 \le m$$

and C_m is the channel capacity of the AWGN channel in state *m* [39],

$$C_m = B \log_2\left(1 + \frac{S}{N}\right) \text{[bits/s]}.$$

where *B* is the bandwidth, *S* is the total signal power and $N = 2\sigma_m$ is the total noise power over the bandwidth in state *m*. The noise variance $\sigma_m^2 = \sigma^2 \frac{m/A + \Gamma}{1 + \Gamma}$, and $\Gamma \equiv {\sigma_G}^2 / {\sigma_I}^2$, $\sigma^2 = {\sigma_G}^2 + {\sigma_I}^2$.

The parameter *A* is the impulsive index, which is given by the average number of received pulses per unit time multiplied by the average pulse width. Furthermore, σ^2 denotes the total noise power and Γ represents the ratio of Gaussian noise power σ_G^2 to impulsive noise power σ_I^2 [19]. An underlying assumption for the Class A model is that the impulses from different interference sources are Poisson distributed in time. The signal to noise ratio (SNR) is defined as $E_s/2\sigma^2$, where E_s is the transmitted signal power. The average capacity, normalized by the bandwidth *B* is then *C*

$$\frac{C}{B}=\frac{1}{B}\sum_{m=0}^{\infty}\pi_m C_m=\sum_{m=0}^{\infty}e^{-A}\frac{A^m}{m!}\log_2\left(1+\frac{E_b}{2\sigma_m^2}\right).$$

4.2 Numerical Results

Numerical results for *C* normalized by the bandwidth as a function of the SNR are shown in Fig. 2 for some different values of *A* and Γ. For large *A* ($A \geq 10$), the class A pdf is very close to a Gaussian distribution, while for *A* and Γ lower than 1 the amplitude pdf will have very heavy tails and the interference can be regarded as very impulsive [21]. By varying the parameters, the pdf can be made arbitrary impulsive or close to a Gaussian distribution. The Middleton's class A model has shown to be a relevant model for many interference sources. For example, a switching-type micro-wave oven has been demonstrated to be modeled well with $A \approx 5 \cdot 10^{-3}$ and $\Gamma \approx 9$ [20].

As can be seen in Figure 5 a less impulsive channel gives a capacity that approaches the capacity for AWGN. Since AWGN is known to be the worst interference with respect to capacity, all values for the different Class A parameter sets give higher values than for AWGN. It should be noted however that the AWGN interference is not always the worst case with respect to the bit error probability, see for instance [23]. We can see in the figure that the capacity for the AWCN channel with the parameter settings (*A*, Γ) = (1, 1), (0.01, 10) and (10, 1) have capacities very similar to the AWGN channel. However, the capacity of AWCN channel with the parameter settings (*A*, Γ) = (0.01, 0.01) and (0.1, 0.1), which are classified as very impulsive interference, differ from the capacity of the AWGN channel.

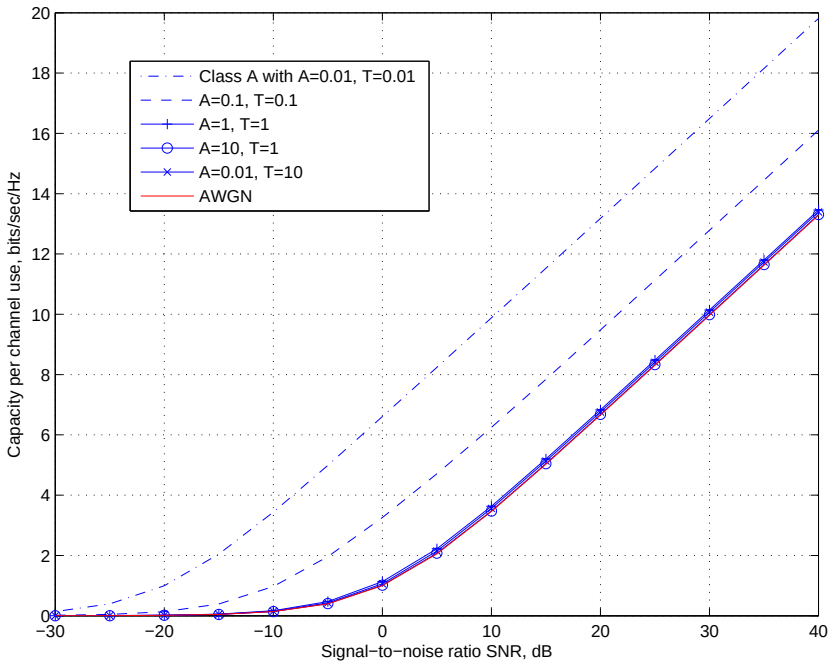

Figure 5: The channel capacity of the AWCN channel for different values of A and Γ.

5 Dynamic Spectrum Allocation

A significant share of recent DSA research concerns the identification of instantaneously un-used "pieces" of spectrum, so-called s*pectrum opportunities*. In particular distributed schemes, where individual or groups of cognitive radios attempt to identify spectrum opportunities that could be used temporarily by secondary users are usually referred to as *overlay spectrum sharing*.

5.1 Spectrum Sensing

In order to use frequency bands dynamically, methods to sense the actual occupancy and interference level in a certain frequency band must be available, both on a higher system level and in some cases in single receivers. A key issue in future dynamic wireless applications is therefore the ability to sense and consider the total electromagnetic interference within the receiver band of the wireless communication system.

Spectrum sensing can be performed on an individual basis or in cooperation with other users. Furthermore, depending on if there is prior information of the users the spectrum sensing can be performed in different ways. If for example the modulation is known for the users that are occupying the spectrum, a matched filter can be used. For users that are using repeatedly transmitted sequences as cyclic prefixes or training sequences, a cyclostationary detection can be performed.

The methods must be fast and of low complexity to be useful in on-line applications and in distributed solutions. Therefore a simple but useful method is tractable to find. Traditional methods for dynamic frequency allocation are based on using pure energy- or power detection only. This means that the energy/power of the interference spectrum is sensed in some way and frequencies with the lowest energy/power are instantaneously chosen for communication.

5.1.1 Interference Temperature

The term "interference temperature" was first proposed by FCC [10]. The idea is that interference is presently minimized through coordination of the frequency, radiated power and location of individual transmitters. Under the interference temperature model, a maximum noise level would be set for an entire band and new systems could be placed in service if it is anticipated that the interference temperature limit would not be exceeded. Several methods are proposed for monitoring interference temperature, but there is no guarantee that any of these theoretical systems will work or that licensed systems will not receive harmful interference.

Several critical points of views have been raised against the idea of interference temperature. One argument against the idea is that since new users would be unlicensed, locating each offender and resolving interference would be an impossible task. Another argument against the idea is that the interference temperature is based on pure power detection only so the interference waveform is not considered. This can for some interference waveforms give large errors [23] in terms of interference impact on a new user. In practical applications, pulsed interference signals cause the largest errors in Bit Error Probability (BEP) when this approach is used. These errors can be in the order of several magnitudes [23], see Figure 6

5.1.2 Spectrum Sensing using an Impulsiveness Correction Factor

The performance of a digital communication system subjected to periodic pulsed interference is analyzed in [25]. In Figure 6 the BEP as a function of the signal-to-interference ratio (*SIR*) is shown for pulse modulated signals with different pulse repetition frequencies (R_S is the symbol rate of the digital communication system). The modulation scheme in Figure 6 is binary phase shift keying (BPSK). The *SIR* is the ratio of the bit energy and the interference power spectral density. The signal-to-noise ratio (*SNR*), defined as the ratio between the bit energy and the power spectral density of the thermal receiver noise, is 12 dB for the calculations shown in the figure. The BEP for the pulsed interference is compared to the BEP for Gaussian noise (additive white Gaussian noise, AWGN). As seen, the BEP for pulsed interference differs significantly from the BEP caused by the AWGN. However, the largest difference in SIR for a constant BEP is 7.5 dB for a shift in pulse repetition frequency f_p with one decade. The analytical proof for this is shown in [26].

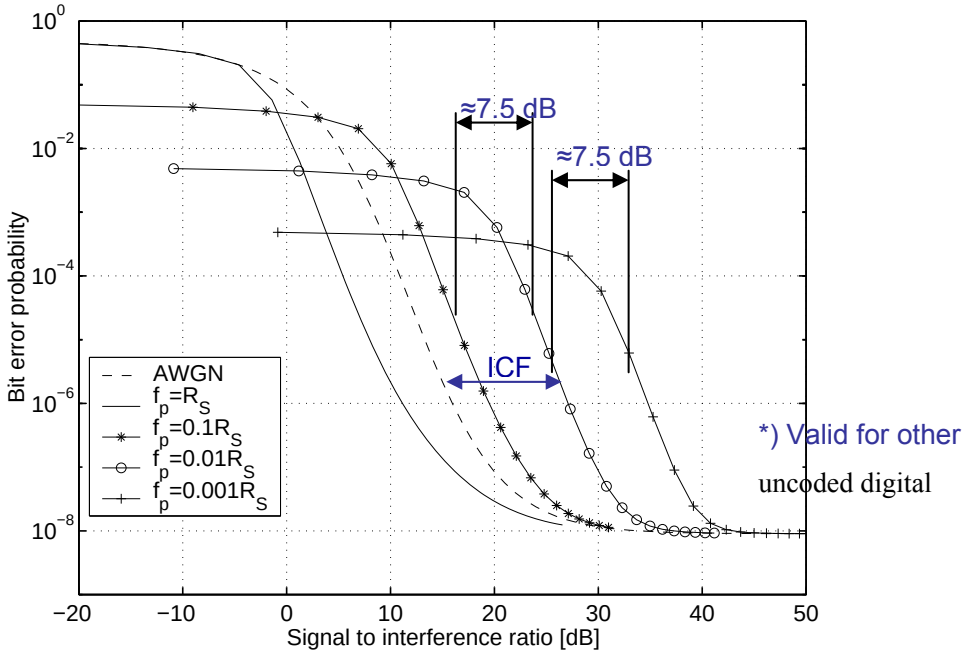

Figure 6. The BEP for pulsed sine wave versus Gaussian (AWGN) interference of a digital communication system.

In [26] it is also shown that this behaviour is true even for other digital modulation schemes. From Figure 6 it is obvious that only using the interference power to predict the impact on a digital communication system can give large errors since the impact is strongly dependent on the waveform properties of the interference signal. These errors can be in the order of several magnitudes or up to a factor 10000 with respect to estimated BEP.

An impulsiveness correction factor (ICF) to adjust for these errors has been proposed in [24]. The ICF can be used as a rough adjustment for the interference-waveform properties so that the measured total interference average power in a frequency band can be used as decision metric in future dynamic applications. The ICF can be used for frequency allocation when the interference impact in terms of the BEP is considered instead of the more simplified way of just

considering the average interference power. We will show that this method can give other recommended frequencies, compared to frequencies selected considering interference power only.

A well-known measure of the impulsive properties of noise is the impulsiveness ratio (*IR*) [27] defined as

$$IR = 20\log\frac{V_{\mathrm{RMS}}}{V_{average}},$$

(1)

where V_{RMS} and V_{average} are the root-mean square and time average values of the envelope of the output of the IF (Intermediate Frequency) filter of a measurement receiver. For periodic pulses with pulse repetition frequency f_p passed through an IF-filter with bandwidth W_{IF}, the *IR* is [28]

$$IR = \frac{\sqrt{W_{\mathrm{IF}}}}{\sqrt{f_{\mathrm{p}}}}.$$

(2)

By inspection of Figure 6, the *ICF* in dB, ICF^{dB}, for BPSK can approximately be expressed as

$$ICF^{\mathrm{dB}} \approx \begin{cases} ICF_{\mathrm{offset}}^{\mathrm{dB}} - 7.5\log\dfrac{f_{\mathrm{p}}}{R_{\mathrm{S}}} & f_{\mathrm{p}} < R_{\mathrm{S}} \\ ICF_{\mathrm{offset}}^{\mathrm{dB}} & f_{\mathrm{p}} \geq R_{\mathrm{S}} \end{cases}, \text{[dB]}$$

(3)

where $ICF_{\mathrm{offset}}^{\mathrm{dB}}$ is a modulation dependent constant which is -4 dB for BPSK. By using the common approximation $W_{\mathrm{IF}} \cong R_{\mathrm{S}}$ we can combine equation (2) and (3) so that

$$ICF \approx -4 + \frac{3}{4}IR \quad , \text{[dB]} \qquad (4)$$

By knowledge of the actual modulation scheme, the corresponding offset is used when the ICF^{dB} is determined. From [23] the $ICF_{\mathrm{offset}}^{\mathrm{dB}}$ for MSK and 64-QAM can be determined to approximately -3 dB and -5 dB respectively. Thus, by knowing the *IR* and the actual modulation scheme of interest, we can determine how much, in terms of *SIR*, the measured interference signal differs from a Gaussian distributed signal causing the same BEP at the victim. Another way of

expressing the application of the *ICF* is that the AWGN approximation for BEP can be used even if the interference signal is not Gaussian. In general, the BEP, P_b ,for AWGN be derived as

$$P_b = f\left(\frac{E_b}{N_0 + N_I}\right), \qquad (5)$$

where E_b is the signal energy per bit [W/Hz], N_0 is the power spectral density [W/Hz] for the receiver noise and N_I is the power spectral density for the interference signal approximated as AWGN within the receiving bandwidth of the wireless receiver of interest. For pulsed interference, the BEP according to (5) can result in errors in the order of several magnitudes, see figure 6. By using the *ICF*, this error can be significantly reduced and restore the usefulness of the AWGN approximation. The corrected BEP, $P_{b,corr}$, can now be denoted as

$$P_{b,corr} = f\left(\frac{E_b}{N_0 + ICF \cdot N_I}\right). \qquad (6)$$

5.1.3 Amplitude Probability Distribution

The Amplitude Probability Distribution (APD) is a statistical function that describes the signal envelope of a user transmitting within a certain frequency band. The APD is defined as the part of time the measured envelope of an interfering signal exceeds a certain level [28]. The relation between the $\mathrm{APD}_R(r)$ and the probability density function of the envelope *R* is

$$\mathrm{APD}_R(r) = 1 - F_R(r) \qquad (1)$$

and

$$f_R(r) = \frac{d}{dr} F_R(r) = -\frac{d}{dr} \mathrm{APD}_R(r), \qquad (2)$$

where $F_R(r)$ and $f_R(r)$ denote the cumulative distribution function (cdf) and probability density function (pdf), respectively. The measurement procedure to measure the APD is standardized [37].

The signal APD is related to the impact this signal will cause another user in terms of bit error probability (BEP). For example, the bound of the BEP for a coherent BPSK receiver subjected to an interference with a measured APD can be interpreted as follows. If the measured bit energy at the detector is E_b, the maximum BEP never becomes higher than $P_{\mathrm{b,max}} = \mathrm{APD}_R(\sqrt{E_\mathrm{b}\frac{Z_0}{T_\mathrm{b}}})$, where Z_0 is the input impedance of the receiver and T_b is the bit time of the radio system [38]. The table below shows the relation between the signal APD and the impact it will generate for at certain radio system with a certain modulation scheme.

Table 1: Bounds derived for different modulation schemes.

Mod.	**Relation $P_{\mathrm{b,max}}$ vs. APD**
2-PSK	$P_{\mathrm{b,max}} \approx \mathrm{APD}_R(\sqrt{E_\mathrm{b}\frac{Z_0}{T_\mathrm{b}}})$
4-PSK	$P_{\mathrm{b,max}} \approx 1/2\,\mathrm{APD}_R(\sqrt{E_\mathrm{b}\frac{Z_0}{T_\mathrm{b}}})$
8-PSK	$P_{\mathrm{b,max}} \approx 1/3\,\mathrm{APD}_R(0.66\sqrt{E_\mathrm{b}\frac{Z_0}{T_\mathrm{b}}})$
16-PSK	$P_{\mathrm{b,max}} \approx 1/4\,\mathrm{APD}_R(0.39\sqrt{E_\mathrm{b}\frac{Z_0}{T_\mathrm{b}}})$
4-PAM	$P_{\mathrm{b,max}} \approx 1/2\,\mathrm{APD}_R(0.63\sqrt{E_\mathrm{b}\frac{Z_0}{T_\mathrm{b}}})$
8-PAM	$P_{\mathrm{b,max}} \approx 1/3\,\mathrm{APD}_R(0.37\sqrt{E_\mathrm{b}\frac{Z_0}{T_\mathrm{b}}})$

16-QAM	$P_{\mathrm{b,max}} \approx 1/4\,\mathrm{APD}_R(0.63\sqrt{E_\mathrm{b}\frac{Z_0}{T_\mathrm{b}}})$
64-QAM	$P_{\mathrm{b,max}} \approx 1/6\,\mathrm{APD}_R(0.38\sqrt{E_\mathrm{b}\frac{Z_0}{T_\mathrm{b}}})$
2-FSK	$P_{\mathrm{b,max}} \approx \mathrm{APD}_R(0.71\sqrt{E_\mathrm{b}\frac{Z_0}{T_\mathrm{b}}})$
4-FSK	$P_{\mathrm{b,max}} \approx 2/3\,\mathrm{APD}_R(\sqrt{E_\mathrm{b}\frac{Z_0}{T_\mathrm{b}}})$

5.2 DSA Considering Interference Waveform Properties

Frequency allocation considering the ICF as measure of the interference impact, of a certain interference average power, can improve performance. In Figure 7, a measured interference spectrum is showed. The spectrum is measured outdoors in the vicinity of a high-voltage installation. The interference has been measured both with an RMS detector and an average detector so that the impulsiveness ratio can be determined. The interference level is then adjusted with the ICF so that an interference level catching the interference impact level can be determined. From this adjusted interference spectrum, the lowest levels are chosen as selected frequencies of use. In Figure 7 frequency lists are shown with the symbols '*' and 'o', where '*' denotes without ICF adjustment and 'o' denotes frequency selected by the ICF adjusted level. In Figure 7 30 frequencies have been selected. As seen the frequency selection will differ if we adjust the interference level with the ICF.

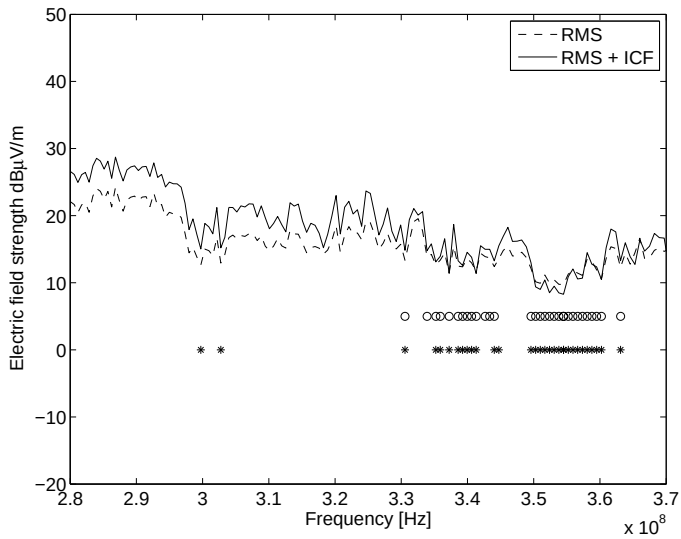

Figure 7: Selected channels both by considering the RMS value of the interference and using the ICF to adjust for waveform properties. '*' denotes without ICF adjustment and 'o' denotes frequency selected by the ICF adjusted level.

In Figure 8 and Figure 9 the distribution of the BEP for the two sets of frequencies are shown. The desired signal level is 19 dBµV/m) and the modulation scheme is BPSK. It can be seen that the number of channel with low BEP is 19 if the ICF is not used. If the ICF is used the number of channels with low BEP is 22 in this case. Thus an overall improvement is achieved by adjusting for the interference waveform properties.

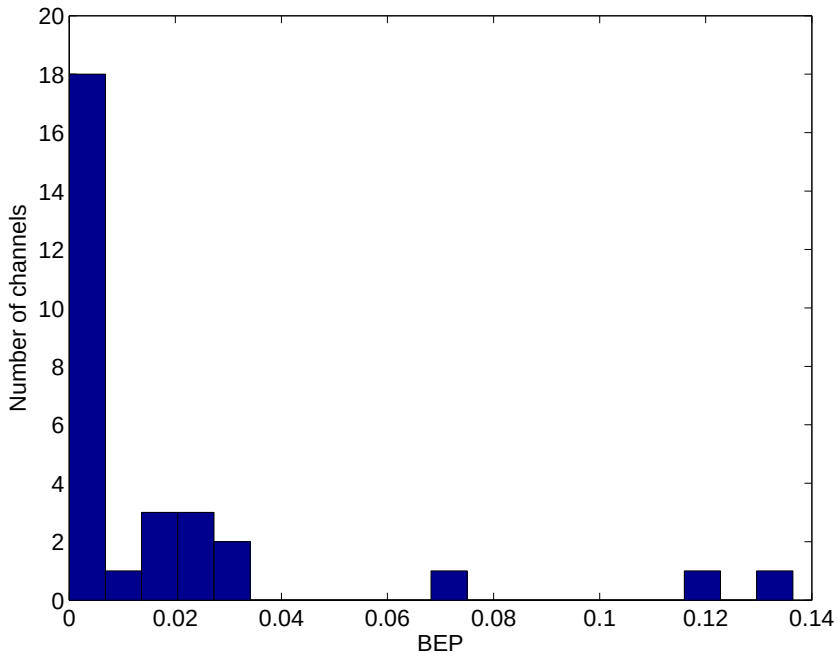

Figure 8: The distribution of BEP over the selected channels if only the RMS value of the interference is considered.

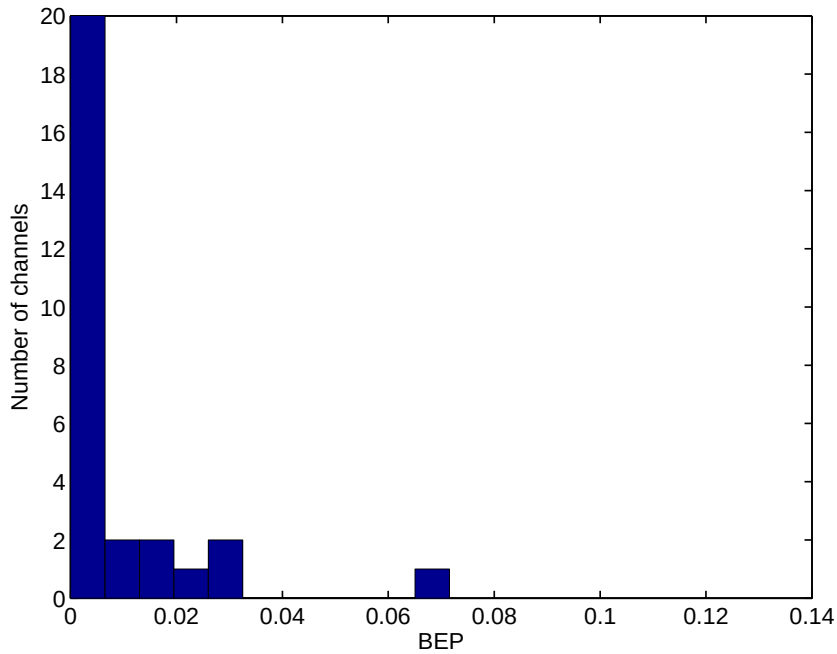

Figure 9: The distribution of BEP over the selected channels if the ICF is used to adjust for interference waveform properties.

5.2.1 Summary of the Proposed DSA Algorithm

The method for dynamic spectrum allocation using the ICF can be summarized in the following steps.

- Measurement of the interference level in the frequency band of interest. Measurement with both rms and average detection.

- Calculate the impulsiveness ratio

$$IR = 20\log\frac{V_{\mathrm{RMS}}}{V_{average}}$$

- Calculate the ICF

$$ICF^{\mathrm{dB}} \approx \begin{cases} ICF_{\mathrm{offset}}^{\mathrm{dB}} - 7.5 \log \dfrac{f_{\mathrm{p}}}{R_{\mathrm{S}}} & f_{\mathrm{p}} < R_{\mathrm{S}} \\ ICF_{\mathrm{offset}}^{\mathrm{dB}} & f_{\mathrm{p}} \geq R_{\mathrm{S}} \end{cases}$$

- Adjust the measured interference spectrum with the ICF

- Select the frequencies with the lowest level after adjustment with the ICF.

6 Conclusions

The Additive White Class-A interference Channel (AWCN) has been proposed as a model for the Cognitive Radio channel. An analytic expression for the channel capacity of the AWCN channel has been derived. Numerical results are shown for some sets of model parameters. For $A > 10$, the capacity of an AWCN channel approaches the capacity of an AWGN channel. Interestingly, as the AWCN channel gets more impulsive, the channel capacity becomes larger than for an AWGN channel.

New DSA algorithms based on the ICF and APD concepts have been derived. The algorithms are shown to give different solutions compared to simple energy detection-based algorithms. The algorithms will be implemented in a demonstrator to evaluate the performance over a real radio channel.

References

[1]. S. Parkvall, E. Dahlman, A. Furuskär, Y. Jading, M. Olsson, S. Wänstedt, and K. Zangi, "LTE-Advanced – Evolving LTE towards IMT-Advanced," in Proc. IEEE Vehicular Technology Conf., Calgary, Canada, Sept. 2008, pp. 1 – 5.

[2]. T. Erpek, M. Lofquist, and K. Patton, "Spectrum occupancy measurements: Loring commerce centre, Limestone, Maine, september 18-20, 2007," 2007, shared Spectrum Company Report.

[3]. First annual report on radio spectrum policy in the European Union; state of implementation and outlook, COM(2004) 507.

[4]. B. Romano, "FCC adopts rule changes for smart radios," 2005, Press Release.

[5]. Joseph Mitola III, "Cognitive Radio – An Integrated Agent Architecture for Software Defined Radio," Ph.D. Thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, May 2000.

[6]. Joseph Mitola III and Gerald Q. McGuire Jr., "Cognitive radio: making software radios more personal," IEEE Personal Communications, vol. 6, issue 4, Aug 1999 Page(s):13 – 18.

[7]. SDR Forum Cognitive Radio Definitions, Working Document SDRF-06-R-0011-V1.0.0, Approved 8 November 2007.

[8]. A. Ghasemi and E. S. Sousa, "Opportunistic spectrum access in fading channels through collaborative sensing," *Journal of Communications*, vol. 2, no. 2, pp. 71–82, Mar. 2007.

[9]. Federal Communications Commission, FCC, Notice of Proposed Rulemaking Regarding "Interference Temperature" Approach for Interference Management (http://www.fcc.gov/sptf/), NPRM 03-289, released November 28, 2003.

[10]. Establishment of an Interference Temperature Metric to Quantify and Manage Interference and to Expand Available Unlicensed Operation in Certain Fixed, Mobile and Satellite Frequency Bands, FCC document ET Docket No. 03-237, August 13, 2004.

[11]. Erik Dahlman, Stefan Parkvall, Johan Sköld and Per Beming, "3G Evolution: HSPA and LTE for Mobile Broadband," Academic Press, 2008.

[12]. IEEE 802 LAN/MAN Standards Committee 802.22 WG on WRANs (Wireless Regional Area Networks), IEEE, http://www.ieee802.org/22/.

[13]. Kia Wiklund, "A new approach for considering the interference impact on digital radio systems from complex interference environments," Chalmers, Sweden, 2007.

[14]. Edgar N. Gilbert, "Capacity of a Burst-Noise Channel," Bell Systems Technical Journal, vol. 39, pp. 1253 – 1265, Sept 1960.

[15]. E. O. Elliot, “Estimates of Error Rates for Codes on Burst-Noise Channels,” Bell Systems Technical Journal, vol. 42, pp. 1977 – 1997, Sept 1963.
[16]. Laveen N. Kanal and A. R. K. Sastry, “Models for Channels with Memory and Their Applications to Error Control,” Proceedings of the IEEE, vol. 66, issue 7, pp. 724 – 744, July 1978.
[17]. REQUEST BY THE EUROPEAN COMMISSION TO THE RADIO SPECTRUM POLICY GROUP FOR AN OPINION ON A COORDINATED EU SPECTRUM POLICY APPROACH CONCERNING WIRELESS ACCESS PLATFORMS FOR ELECTRONIC COMMUNICATIONS SERVICES (WAPECS)”, Brussels, 26 May 2004. DG INFSO/B4, RSPG04-45 Rev.
[18]. “Second report and order and memorandum opinion and order”, FCC 08-260, November 14, 2008.
[19]. D. Middleton, “Canonical and quasi-canonical probability models of class A disturbance,” IEEE Trans. on EMC, vol. 25, pp. 76–106, May 1983.
[20]. J. Häring and A. J. Han Vinck, “Performance bounds for optimum and suboptimum reception under class-A impulsive noise,” IEEE Transactions on Communications, vol. 50, pp. 1130–1136, July 2002.
[21]. A. D. Spaulding and D. Middleton, “Optimum reception in an impulsive interference environment-part I: Coherent detection,” IEEE Transactions on Communications, vol. 25, pp. 910–923, September 1977.
[22]. A. J. Goldsmith and P. P. Varaiya, “Capacity of fading channels with channel side information,” IEEE Transactions on Information Theory, vol. 43, pp. 1986–1992, November 1997.
[23]. P. F. Stenumgaard, “A simple impulsiveness correction factor for control of electromagnetic interference in dynamic wireless applications,” IEEE Communications Letters, vol. 10, pp. 147–149, March 2006.
[24]. P. Stenumgaard and K. Wiklundh, “Dynamic Frequency Allocation Considering Interference Waveform Properties”, *submitted to DYSPAN 2010.*
[25]. P. F. Stenumgaard, “On radiated emission limits for pulsed interference to protect modern digital wireless communication systems”, *IEEE Transactions on Electromagnetic Compatibility* vol. 49, no. 4, November 2007.
[26]. Peter F. Stenumgaard, “Using the Root-Mean-Square detector for weighting of disturbances according to its effect on digital communication services” *IEEE Transactions on Electromagnetic Compatibility*, vol. 42, pp. 368-375, November 2000.
[27]. ITU Document “The Protection of Safety Services from Unwanted Emissions”, AMCP WGF/6 WP/6, Document 1/13-E, 30 October 2000.
[28]. IEC CISPR 16-1/1999-10: Specification of radio disturbance and immunity measuring apparatus and methods. Part1: radio *disturbance and immunity measuring apparatus.*

[29]. B. Le (2007).'Building a Cognitive Radio – From Architecture Definition to Prototype implementation'. In Electrical and Computer Engineering. Vol. Doctor of Philosophy: Virgina Polytechnic Institue and State University
[30]. L. E. Doyle, *Essentials of Cognitive Radio*, Cambridge, 2009.
[31]. S. Couturier & B. Scheers, "The State of the Art of Dynamic Spectrum Access," Military Communications and Information Systems Conference (MCC) 2009.
[32]. E. Hossain, D. Niyato, Z. Han, *Dynamic Spectrum Access and Management in Cognitive Radio Networks*, Cambridge, 2009.
[33]. J. Neel, R. M. Buehrer, J. H. Reed, R. P. Gilles (2002), "Game Theoretic analysis of a Network of Cognitive Radios," Circuits and Systems, 2002. MWSCAS-2002. The 2002 45th Midwest Symposium on Volume 3, 4-7 Aug. 2002 Page(s):III-409 - III-412 vol.3
[34]. G. Yue, "Antijamming Coding Techniques," IEEE Signal Processing Mag., Vol. 25, No. 6, Nov. 2008, pp 35 – 45.
[35]. J. J. Boutros, A. Guillen i Fabregas, E. Biglieri and G. Zemor, "Low-Density Parity-Check Codes for Nonergodic Block-Fading Channels," available online at http://arxiv.org/abs/0710.1182
[36]. K. C. Wiklundh, P. F. Stenumgaard, and H. M. Tullberg, "Channel capacity of Middleton's class A interference channel," IET Electronics Letters, Vol. 45, No. 24, Nov 2009, pp. 1227 – 1229.
[37]. CISPR 16-1-1, ed. 2:2006-03.
[38]. K. Wiklundh, "Relation between the amplitude probability distribution of an interfering signal and its impact on digital radio receivers," IEEE Trans. On EMC, Vol. 48, No. 3, Aug. 2006, pp. 537-544.
[39]. C. E. Shannon, "A mathematical theory of communication," The Bell System Technical Journal, vol. 27, pp. 379–423, 623–656, July / October 1948.

Part 2

7 Security

7.1 Security Architecture

The main focus of the project DynamIT is research around methods for a cognitive radio, CR, to perceive the environment, e.g. utilization of frequencies, and to use this information to adjust parameters of the radio. The perception of the environment might be done autonomously by a radio node, or it might be a joint task of many radio nodes, helping each other by exchanging data. In either case, there are security issues, since the perception could possibly be disturbed by a hostile opponent. Therefore, it is important to have security measures, intrinsic in the cognitive radio node, to mitigate such attacks. The structure of this intrinsic security is dependent on characteristics of the perception techniques, however.

In addition to the named intrinsic security, there are also security requirements on the architecture of the system. A radio node primarily acts as an intermediary, transmitting information from an information source towards a final receiver. Thus, the radio node must be trusted to be consistent with the security policy of the system as a whole. It must be able to ensure the integrity of the transmitted information, and it must be able to prevent unauthorized access to information. This security functionality could be coined as provided security, as opposed to the intrinsic security. The requirements are dependent on what types of information that should be transmitted. The requirements on architecture and security are extensive for a node trusted for classified information.

An overview of research areas of interest for the project DynamIT was presented in memorandum [memo08]. Security aspects for, SDR, military software defined radios, notably American JTRS and Swedish GTRS, were discussed. These follow the SCA [15], software communications architecture. The security architecture can, very roughly, be depicted as Figure 1. The figure is copied from wikipedia [16], but the same kind of figure can be found in many papers. The central security aspect is that the radio node is divided into two parts, red side and black side, respectively. The only way to transmit information shall mandatorily be via a crypto subsystem, allegedly assuring that all information in the radio (black side) is always protected by appropriate encryption. However, the figure is too rough. In reality much information, e.g. control signals between red and black, must bypass the crypto subsystem. If the security policy is demanding, e.g. separation of information types, and if all functionality is realized in one system, the security solution in the system will be hard to assure.

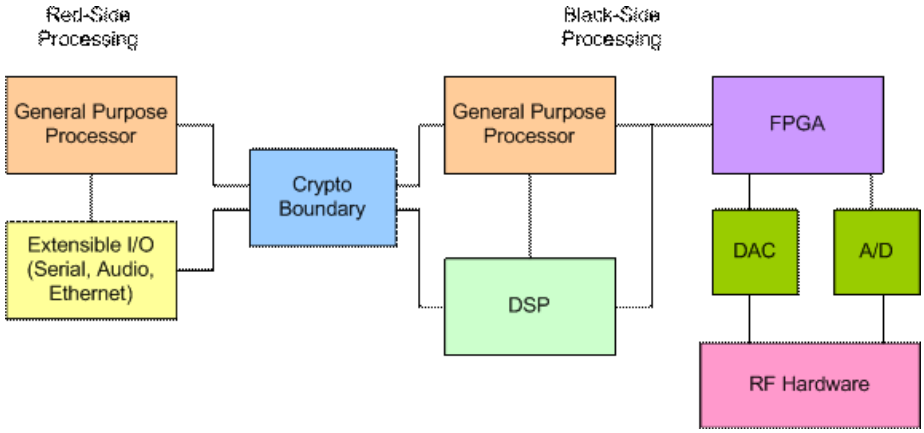

Figure 10 Main architecture of a military SDR.

In [18] areas for security research are listed. They are summarized as:

1. Separation. The radio must be capable of keeping information separate. The requirements are essentially the same as for access control in a general multi-user computer system.
   Particular attention must be paid to handling of information at different levels of security classification. Mandatory separation of classification levels is often required. The separation red/black might be best achieved by running the crypto subsystem in a virtual machine, which is a very important research issue.
2. COMSEC. Cryptography is not a research area within FOI. But one question is if the system should provide link-link encryption, network end-end encryption, or both, and how the encryption should be supervised. This is related to 1. Separation.
3. Proxy. The radios act as intermediary proxies. The encryption, and other security functions in the radio, must work together with the end systems. The end systems might for instance define classification levels, priority levels etc. This is related to 2.
4. Protocol. The radios form a transport network. Most likely, a special protocol is needed for control of the communication, above all for CR. Since the communication is by radio, the security issues are vital. Authentication and management of cryptographic keys are particularly important. An efficacious authentication method is essential.

5. Downloads. SDR means a possibility to change configuration by software downloads. This is a variant of 4, with excessive security demands. The same goes for rekeying over the air.

Points 4–5 are relevant to the intrinsic security. All points 1–5 are relevant to the provided security. We have decided to begin with points 1 and 3, since they are not dependent on details of chosen techniques for cognitive perception.

Our chosen security architecture can be depicted as Figure 11. It is a variant of Figure 10. The most significant difference to Figure 10 is the extensive use of virtual machines. The rationale for this is elaborated in the sections further down. It is noteworthy, that the architecture results in a possibility to implement parts as physical machines instead of virtual machines. For instance, one crypto subsystem might be in a separate physical machine, while other cryptosystems might be virtual machines. The architecture is inspired by Davidson, MILCOM 2008 [14].

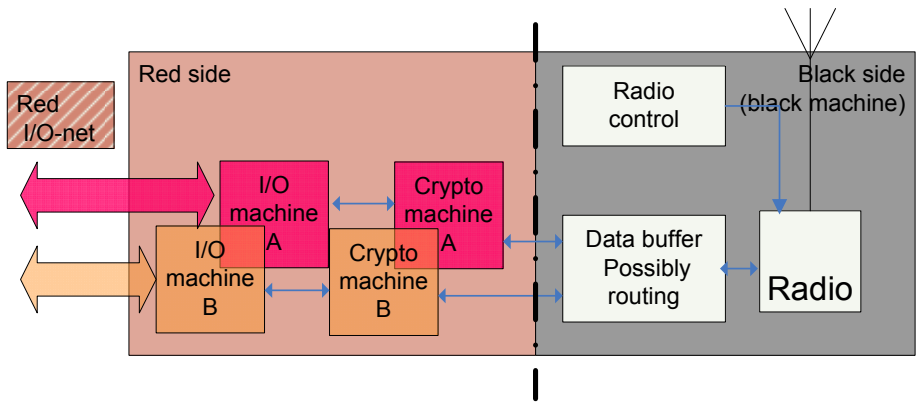

Figure 11 Security architecture with virtual machines.

The main motivation to the depicted two parallel I/O-tracks is to facilitate separate handling of information classes. The two classes could be for instance different priority classes, or they could be different security classes. The latter example calls for a high assurance level of the system. This is alleviated by the use of virtual machines as opposed to a single, complicated system.

The conclusion in Section 7.3 is that mandatory control of access between virtual machines facilitates the separation of information classes, A and B respectively. A subtle spot in the figure is the data buffer at the black side. Packets received by the radio shall be routed to the correct crypto machine. In case the routing fails,

the information would not be accessible due to the encryption anyway. But the routing of received packets still should be carefully devised.

7.2 Separation Kernel and Multiple Independent Levels of Security and Safety, MILS

For a complex system, e.g. an SDR like Figure 10, it is hard to reach a high assurance level when security matters shall be evaluated. A natural alternative is to find ways to divide the system into a set of smaller partitions, which would be less difficult to assure. However, the partitions must be isolated and independent, otherwise the complexity would be increased instead of reduced.

This was first discussed in 1981 by Rushby [1][2], who coined the term Separation Kernel for a small operating system which enforced the separation. The concept has been further developed in the safety critical community, e.g. the avionics community. It is described for instance in [8], where the term MILS, Multiple Independent Levels of Security and Safety, is used. The most important security qualities in the separation kernel for MILS are stated as:

- Data Separation. The memory address spaces of a partition must be completely independent of other partitions.
- Information Flow. Partitions must communicate with each other. For secure systems, authorized communication channels between partitions must be defined.
- Sanitization. The separation kernel is responsible for cleaning any shared resources (microprocessor registers, system buffers, etc.) before a process in a new partition can use them.
- Damage Limitation. Address spaces of partitions are separate, so an errant process in one partition can not affect processes in other partitions. The separation kernel will also enforce bounds on shared resources, providing guaranteed minimum processing time, memory and other resources to the partitions.

The properties of separation and isolation are in the same way important to sustain security as well as safety. The security requirements on a separation kernel are stated by USA NSA/CSS[4] in a protection profile for high robustness [3]. A protection profile contains evaluated requirements, both on security functions and on assurance methods. High robustness is the highest assurance level for systems containing classified information. NSA/CSS has defined three

[4] National Security Agency/Central Security Service

assurance levels – basic, medium and high. There is one evaluated product [4], from Green Hills Software [5] conformant to the high robustness profile. According to Green Hills' web page, the product has been used as a basis for an SDR. However, it is not suitable as a laboratory research platform. Instead, we intend to use a virtual machine monitor as a basis for the architecture in Figure 11.

7.3 Virtual Machines

Virtual machines is the term used when one host machine emulates many guest machines. The Virtual Machine Monitor, VMM, on the host could be a full fledged operating system, like Linux or MS Windows. Alternatively, it could be a much smaller, by a factor 100 – 1000, special operating system, often called hypervisor. A hypervisor has many of the qualities in a separation kernel. The security qualities of virtual machines were reported in [17]. Some abstracts are made in the following.

IEEE Security&Privacy has a special issue Sep/Oct 2008, "Virtualization and Security: Back to the Future" [7]. The title alludes to the fact that virtualization is an old concept. The Figure 12 Figure 14 of virtual machines are copied from [7], through the courtesy of IEEE.

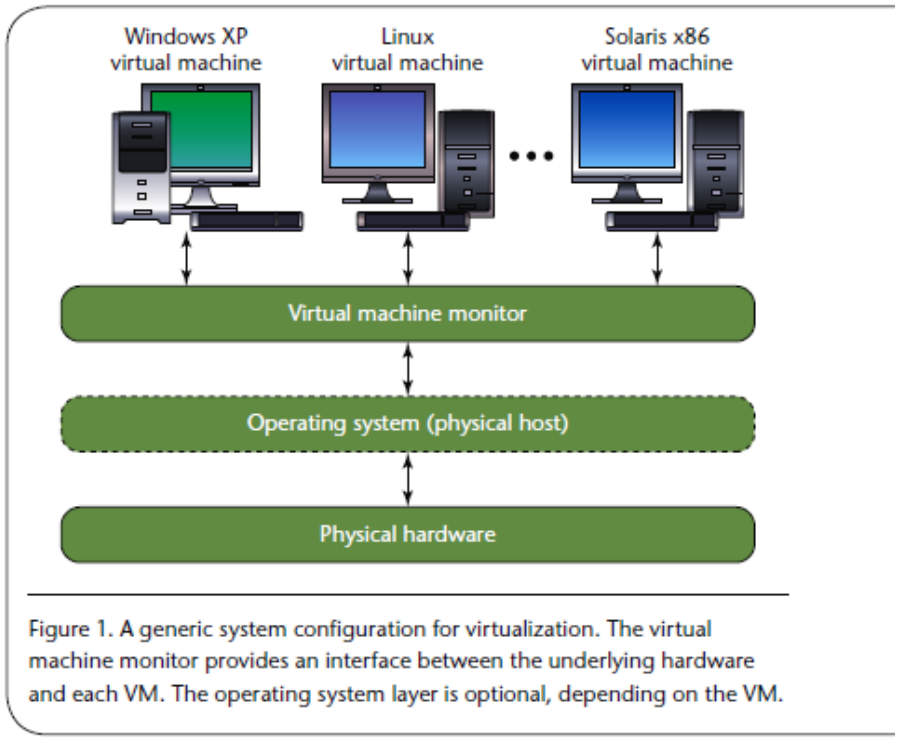

Figure 12 A generic system configuration for virtualization, from [13], © 2008 IEEE.

The optional host (dotted line) means that the VMM could be an “ordinary” application running on an “ordinary” host OS. Virtualization as a means to avoid piles of physical computers in computer centers often looks like this.

For isolation purposes, like MILS [8], Multiple Independent Levels of Security, it is preferable to have a small (which hopefully could be assured) hypervisor running directly at the hardware, i. e. no big host OS and no dotted box in Figure 12. Essentially this means that the hypervisor is a “basic OS”, below the “application OS”. The VMs are machines, including an operating system (preferably stripped to a minimum), running on top of the hypervisor. Small size, both for hypervisor and stripped VM, is a good thing from a security point of view.

Such a hypervisor can essentially contain the security qualities required for a separation kernel. This is discussed in [17], particularly for the open source hypervisor Xen [6][10]. Xen is sketched in Figure 13.

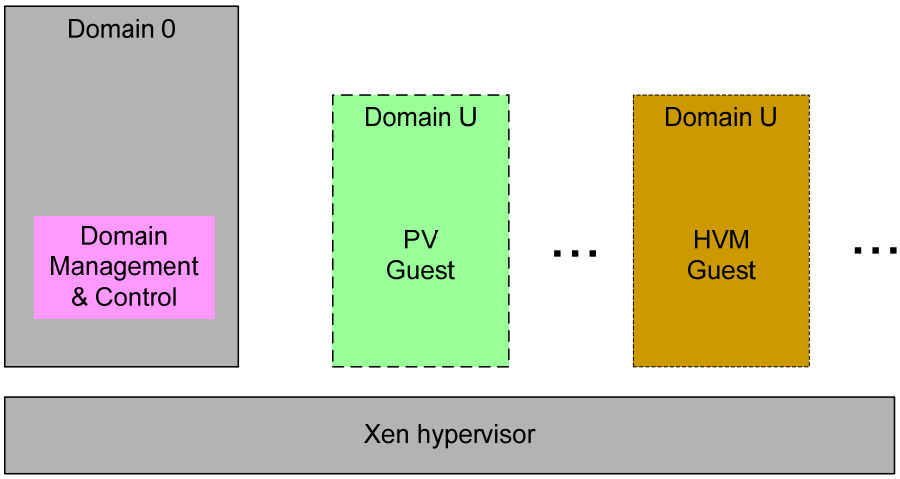

Figure 13 Basic organization of Xen.

In Xen the separated partitions are called domains. Domains U (for user) run the guest machines, with (stripped) ordinary OSs like Linux or MS Windows. A guest OS could be modified (not MS Windows), called paravirtualized, PV, which means better performance, particularly for I/O. Alternatively, it could be unmodified (HVM[5]), which is the only option for MS Windows. Domain 0 runs in "half privileged" mode, at a privilege level between hypervisor and user. It often contains I/O-drivers, called by modified PV guests. Since domain 0 has higher privilege level than user, it is part of what commonly is called TCB, Trusted Computing Base, and its security has to be assured.

The required quality "*Information Flow*" is rather primitive in basic Xen. But extensions exist, which allow policies for authorized access control via a monitored channel between a pair of VMs. An example is sHype [12], with an origin from IBM. For details, see [17].

[5] The surprising acronym HVM comes from Hardware Virtualized Machine

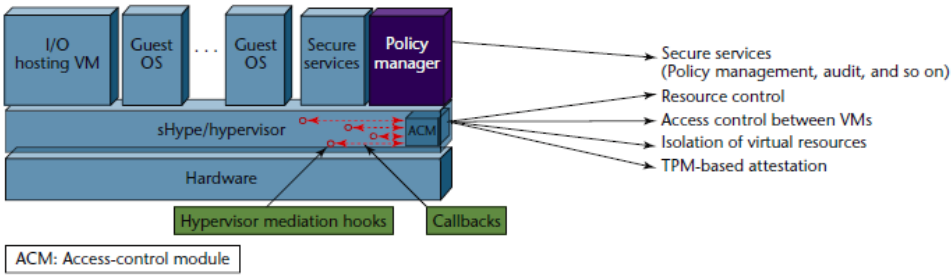

Figure 14: sHype hypervisor security architecture, from [9], © 2008 IEEE.

The sHype extension, Figure 14, adds hooks to the hypervisor at places where the communication channel is handled. The hooks call a policy manager, which runs in a special “half privileged” machine. Our belief is that this results in an authorized and mandatory access control, which is useful in the proposed architecture in Figure 11.

References

[1] "Design and Verification of Secure Systems", Reprint of a paper presented at the 8th ACM Symposium on Operating System Principles, Pacific Grove, California, 14–16 December 1981. (ACM Operating Systems Review Vol. 15 No. 5 pp. 12-21), http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.62.1726&rep=rep1&type=pdf

[2] B Randell, J Rushby, "Distributed secure systems: Then and now", Proceedings of the Twenty-Third Annual Computer SecurityApplications Conference, pages 177–198, IEEE Computer Society, Miami Beach, FL, December 2007. Invited "Classic Paper" presentation, http://www.cs.newcastle.ac.uk/publications/trs/papers/1052.pdf

[3] Information Assurance Directorate, National Security Agency, Fort George G. Meade, MD, 20755-6000. U.S. Government Protection Profile for Separation Kernels in Environments Requiring High Robustness, June 2007. Version 1.03.

[4] Green Hills Software, INTEGRITY-178B Separation Kernel, http://www.niap-ccevs.org/cc-scheme/st/vid10119/

[5] Green Hills Software Inc, http://www.ghs.com/

[6] Xen Hypervisor, http://www.xen.org/

[7] Virtualization and Security: Back to the Future, IEEE Security&Privacy, 2008, volume 6, Issue 5, http://ieeexplore.ieee.org/xpl/tocresult.jsp?isYear=2008&isnumber=4639007&Submit32=View+Contents

[8] J. Alves-Foss, P. W. Oman, C. Taylor, W. S. Harrison, "The MILS architecture for high-assurance embedded systems", International Journal of Embedded Systems, Volume 2, Number 3-4 / 2006, pp 239-247, http://inderscience.metapress.com/app/home/contribution.asp?referrer=parent&backto=issue,9,10;journal,6,9;linkingpublicationresults,1:110847,1

[9] R. Perez, R. Sailer, L. van Doorn, Virtualization and Hardware-Based Security, IEEE Security&Privacy, 2008, volume 6, Issue 5, pp 24-30

[10] Xen Architecture Overview, http://wiki.xensource.com/xenwiki/XenArchitecture?action=AttachFile&do=get&target=Xen+Architecture_Q1+2008.pdf

[11] R. Sailer, T. Jaeger, E. Valdez, R. Caceres, R. Perez, S. Berger, J. L. Griffin, L. var Doorn, “sHype: Mandatory Access Control For XEN”, http://www.docstoc.com/docs/5691508/sHype-Mandatory-Access-Control-For-XEN

[12] R. Sailer , T. Jaeger , E. Valdez , R. Caceres , R. Perez , S. Berger , J. L. Griffin , L. van Doorn, “Building a MAC-Based Security Architecture for the Xen Open-Source Hypervisor”, Proceedings of the 21st Annual Computer Security Applications Conference, p.276-285, December 05-09, 2005 [doi>10.1109/CSAC.2005.13], http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=1565255&isnumber=33214

[13] P. A. Karger, D. R. Safford, I/O for Virtual Machine Monitors – Security and Performance Issues, IEEE Security&Privacy, 2008, volume 6, Issue 5, pp 16-23

[14] J. A. Davidson, “On the architecture of secure software defined radios”, Military Communications Conference, 2008. MILCOM 2008. IEEE

[15] http://sca.jpeojtrs.mil/

[16] http://en.wikipedia.org/wiki/Software_Communications_Architecture

[17] A. Bengtsson, L. Westerdahl, “Virtual Machines, Security Qualities”, FOI-R--2904--SE, December 2009

[18] H. Tullberg, P. Stenumgaard, A. Bengtsson, M. Sparf, “Dynamic Telecommunications – Overview of Research Problems”, FOI Memo 2473, June 2008