HATE SPEECH ON SOCIAL NETWORK SITES: PERPETRATOR AND SERVICE PROVIDERS' LIABILITY

Frans Marx
BCom Bluris LLB LLD
Professor, Private Law
Nelson Mandela Metropolitan University
Port Elizabeth

SUMMARY

The article investigates the phenomenon of hate speech on social network sites and gives an overview of the national and international legal instruments which are available to combat hate speech. After an overview of the nature of hate speech and the early international attempts to curb it, hate speech in South Africa is investigated. The question is posed whether statements of hatred made on the Internet, especially if published from sites such as Facebook which is external to South Africa, can lead to liability for perpetrators in South Africa. International responses to hate speech in cyberspace are then investigated with specific reference to the possible liability of Internet service providers for hate speech posted by third parties on their websites. It is shown that, although service providers in the United States enjoy more protection than those in European Union, Canada and South Africa, hate speech on social network sites can be legally curbed. It is concluded that the myth that the Internet as a godless, lawless zone can and must be dismissed.

1 INTRODUCTION

Recently *News24* reported under the heading, "Malema fans count dead 'boere'"[1] that a Malema fan, Clearence Letlonkane, has been posting information on Facebook about farmers being murdered. One of the statements read: "3000 farmers dead since '94 … we're far from being even … So kill da boer, kill da farmer." When asked about these and other statements, by the News24 reporter, Letlonkane commented: "I haven't killed anyone, but I am not sympathetic to those who have been murdered … sue me for not shedding a tear." This statement on Facebook is but one example of hate speech found on the Internet. In fact, the Internet is

[1] Du Plooy *News24* 17 March 2010 http:/www.news24.com/SouthAfrica/News/Malema-fan-counts-dead-boere-20100316 (accessed April 2010).

proliferated with so-called hate sites, the sole purpose of which is to spread hate and to propagate hatred against different groups.[2]

If the statement mentioned above was made in a public forum in South Africa, it probably would have been regarded as hate speech and therefore unprotected speech in terms of section 16(2) of the Constitution of the Republic of South Africa, 1996 (hereinafter "the Constitution"). It might even have been considered as prohibited speech as envisaged by section 10 of the Promotion of Equality and Prevention of Unfair Discrimination Act 4 of 2000. However, since this statement was made on the social network site, Facebook, the question arises what legal consequences, if any, can such statement, if published in cyberspace, have.

Since the advent of the Internet, states have struggled to regulate it. As Johnson and Post stated in 2000:[3]

> "The non-geographical character of the net makes it very difficult to apply current, territorially-based rules to activities online ... Local sovereigns may have a monopoly on the lawful use of physical force, but they cannot control online actions whose physical location is irrelevant or cannot even be established."

Similarly Akdeniz stated in 2007:[4]

> "The global, decentralized and borderless nature of the Internet creates a potentially infinite and unbreakable communications complex which cannot be readily bounded by one national government or even several or many acting in concert; there is simply no unique solution for effective regulation at the national level."

Statements like the ones mentioned above, create the impression that cyberspace cannot be subjected to the rules of the "real" world and cannot therefore be legally controlled. The fact is: the Internet (cyberspace), just like any other entity, is subject to state laws and most states have special legislation pertaining to it. In South Africa many, but not all aspects of the Internet, are governed by the Electronic Communications and Transactions Act 25 of 2002. Other legislation, for example, the Films and Publications Act 65 of 1996 also contains provisions prohibiting certain actions on the Internet.[5] Cyberspace, like any other space within the borders of the country, is subject to the laws of the country. The problem is that laws govern territories and cyberspace transcends territories.

---

2 See Roos "Freedom of Expression" in Van der Merwe, Roos, Pistorius, Eiselen *Information and Communications Technology Law* (2009) 445. According to Center "Online Hate Sites Grow with Social Networks" 16 March 2010 *The New York Times*, hate sites increased as a result of social networking and at the time of writing over 11 500 hate sites existed.

3 Johnson and Post "And How Shall the Internet be Governed? A Meditation on the Relative Virtues of Decentralized Emergent Law" http://www.cli.org/emdraft.html quoted in UNESCO Publishing *The International Dimensions of Cyberspace Law* (2000) 35.

4 Akdeniz "Governing Racist Content on the Internet: National and International Responses" 2007 56 *University of New Brunswick Law Journal* 103 109.

5 *Eg*, s 24(2) of the Act criminalizes the distribution or possession of child pornography.

A further problem, often stated, is that the law always lags behind technology.[6] As far as the Internet is concerned, the advent of Web 2 and the explosion of social network sites such as Facebook, Twitter and MySpace, where third parties upload content seemingly beyond the control of the service providers, may illustrate the point. Chapter XI of the Electronic Communications and Transactions Act deals with service providers' liability. Certain categories of service providers are, in terms of this chapter, exempted from liability for content posted by third parties, provided that they fulfil certain requirements. By nature, Web 2.0 social network sites differ from other Internet service providers in that they are almost completely based on user-generated content.[7] These sites encourage their users to upload and maintain their own content and they give users almost unfettered control over such content without interference from the providers of the service. This makes effective legal control of these sites even more problematic.

The purpose of this article is to investigate whether statements such as the one mentioned in the introduction to this article, if made on social network sites, can qualify as hate speech as envisaged by section 16(2) of the Constitution or section 10 of the Promotion of Equality and Prevention of Unfair Discrimination Act 4 of 2000 and, whether there are remedies in South African law, either against the person making such a statement on a social network site such as Facebook, MySpace, Twitter and the like, or against the social network site itself. This will entail an investigation into the local, national and international regulation of hate speech in cyberspace. The article will therefore, firstly investigate the nature of hate speech, then look at international responses to hate speech, and then investigate the South African position relating to hate speech. This will be followed by an overview of international instruments dealing with the regulation of hate speech in cyberspace and an overview of how some states deal with the phenomenon. Network neutrality and intermediate liability of service providers will be discussed with reference to the safe-harbour provisions in the US, the United Kingdom and South Africa. Finally, the possibility of trans-border remedies will be investigated.

It must be appreciated that the above topic covers a wide field where many legal rules are not fully developed and where there is still much legal uncertainty. Thus, due to the broad nature of the topic and space constraints, this article only attempts to give an overview of the most important rules and principles governing this field. It is hoped that further research and critical analysis of many of the subtopics will be forthcoming.

---

6 Moses "Recurring Dilemmas: The Law's Race to Keep Up with Technological Change" 2007 *University of Illinois Journal of Law, Technology and Policy* 239.

7 *Brown* "Fortifying the Safe Harbours: Re-evaluating the DCMA in a Web 2.0 World" 2008 23 *Berkeley Technology LJ* 437.

2 THE NATURE OF HATE SPEECH

Hate speech as a phenomenon received attention from the international community after the atrocities of the Second World War and more specifically as a result of the holocaust and the persecution of Jews and other ethnic minorities. It is not surprising that the early regulation of hate speech was directed at the prohibition of racial discrimination. The first international agreements aimed at the prevention of discrimination were concluded in Europe during the 1960s.[8] Most notable is the International Convention on the Elimination of All Forms of Racial Discrimination[9] which was signed in 1966.[10] Subsequently many states passed laws to prevent discriminatory speech and in time the regulation of more than just discrimination on grounds of race was targeted. In many cases protection was extended to other vulnerable groups such women, religious groups and minority groups with specific characteristics.

Hate speech as an international phenomenon is not easy to define. According to Bacircioglu[11] "it covers abusive denigrating harassing speech targeting a group's or individual's national, racial, religious or ethnic identity. Yet there is no universally acknowledged definition. According to Human Rights Watch hate speech is 'any form of expression regarded as offensive to racial, ethnic and religious groups and other discrete minorities, and to women'. Some scholars define it as a 'generic term that has come to embrace the use of speech attacks based on race, ethnicity, religion, and sexual orientation or preference'." According to Roos,[12] hate speech "is generally understood as meaning epithets or disparaging and abusive words and phrases directed at individuals or groups representing a specific race, religion, ethnic background, gender or sexual preference".

If one analyses these definitions, it soon becomes clear that hate speech involves an utterance that is derogatory, abusive or offensive; it is directed at an individual or group; and it is based on the fact that the subject or subjects belong to a specific group, be it an ethnic group, a group belonging to a specific race or a group that is identifiable because of one other common characteristic, such as gender, sexual orientation, religion, *etcetera*. Racist

---

8 Blarcum "Internet Hate Speech: The European Framework and the Emerging American Haven" 2005 62 *Washington & Lee LR* 781 785.

9 This convention was signed and ratified by the US but the US made a reservation indicating its refusal to undertake any measures that violate the First Amendment. See Blarcum 2005 62 *Washington & Lee LR* 786.

10 International Convention on the Elimination of All Forms of Racial Discrimination 1966, 660 U.N.T.S. 195, http://www2.ohchr.org/english/law/cerd.htm (accessed 2010-07-20).

11 Bakircioglu "Freedom of Expression and Hate Speech" 2008-2009 16 *Tulane Journal of Comparative and International Law* 1 4; and see also Hee Han "Hate Crimes and Hate Speech" 2006 7 *The Georgetown Journal of Gender and the Law* 679 680: "Hate speech is bias motivated speech aimed at a victim because of some characteristic or characteristics of the victim. The perpetrator, because of these characteristics, identifies the victim as a member of a particular social group based on gender, sexual orientation, race, ethnicity, religion, national origin or disability in order to insult or to offend the victim or to establish a threat, libel or fighting words in a derogatory or demeaning manner."

12 444.

speech (prevented as such in many states) is therefore a category of hate speech. It is not surprising that, as will be shown below, many states have legislation which prohibit either hate speech or racist and xenophobic speech.

3 INTERNATIONAL RESPONSES TO HATE SPEECH

In 1966, long before the advent of the Internet, the United Nations promulgated the International Covenant on Civil and Political Rights in 1966.[13] Article 20(2) provides that "any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law". This obligation was echoed in the UN Convention on the Elimination of all Forms of Racial Discrimination which has been in force since 1969.[14] In terms of the Convention, parties "shall declare an offence punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination".[15] Article 4 prohibits the advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence. By 2010 the International Convention on the Elimination of All Forms of Racial Discrimination was ratified by 173 states.[16] Following the Convention, almost all European states and many other states proscribe hate, racist or discriminatory speech in one or other form.[17]

In the United Kingdom for instance, freedom of expression is guaranteed in the Human Rights Act of 1998. The United Kingdom has, however, acknowledged that racial hatred can disturb the public order. Speech that causes racial hatred is prohibited in Part III of the Public Order Act of 1986.[18] Conduct which is threatening, abusive or insulting and is intended to or is likely to stir up racial hatred, is criminalized. If an offence is committed in terms of the Public Order Act with the consent or connivance of a director, manager, secretary, member or similar officer of a corporation, both the offending party and the corporation are guilty of the offence and can be prosecuted.[19]

---

13 International Covenant on Civil and Political Rights, G.A. Res. 2200A [XXI] U.N. Doc A/6316 (Dec. 16, 1966).

14 See Rorive "What Can be Done Against Cyber Hate? Freedom of Speech versus Hate in the Council of Europe" 2009 17 *Cardozo Journal of Comparative and International Law* 417 419.

15 International Convention on the Elimination of all Forms of Racial Discrimination, G.A. Res. 2106 [XX] art. 4a, U.N. Doc A/6014 (Dec. 21, 1965).

16 http://treaties.un.org/Pages/Treaties.aspx?id=4&subid=A&lang=en (accessed 2010-10-11).

17 Roos 445.

18 Burns *Communications Law* (2009) 154; *Legal Instruments for Combating Racism on the Internet*, a report prepared and updated by the Swiss Institute of Comparative Law (Lausanne); Council of Europe Publication Strasbourg 2009 95; and Cucereanu *Aspects of Regulating Freedom of Expression on the Internet* (2008) 35.

19 In terms of s 28-29 of the Public Order Act of 1986. A guilty party can receive a sentence of up to seven years, a fine or both in terms of s 27(3) and s 29 of the Public Order Act of 1986.

In the German *Strafgesetzbuch* attacks on human dignity which are likely to breach peace, are criminalized.[20] Hate speech is also proscribed in many states outside Europe. In Canada for instance, speech that incites hate against an identifiable group and which is likely to lead to breach of peace, is a criminal offence in terms of the Canadian Criminal Code[21] In the United States, because of the prevalence of the First Amendment to the American Constitution regarding freedom of expression, hate speech *per se* is not prohibited but so-called fighting words, that is words that incite imminent violence, are prohibited.[22]

4 HATE SPEECH IN SOUTH AFRICA

In the South African context, section 16(1) of the Constitution provides that everyone has the right to freedom of expression, including, *inter alia*, freedom of the press and the media. Section 16(2) provides that the right to freedom of expression does not extend to "(a) propaganda for war; (b) incitement of imminent violence; or (c) advocacy of hatred that is based on race, ethnicity, gender or religion, and that constitutes incitement to cause harm".

In *Freedom Front v SA Human Rights Commission*,[23] the SA Human Rights Commission had the opportunity to decide whether the slogan "Kill the Farmer, kill the Boer", chanted at high-profile functions organized by the African National Congress, amounted to hate speech in terms of section 16(2)(c). Commissioner Govender pointed out that, unlike the jurisprudence developed around freedom of expression in the United States where the First Amendment is deemed to be the pre-eminent right, South African courts have adopted a much more nuanced and balanced approach. The right to freedom of expression in section 16 of the Constitution is not pre-eminent and the values of human dignity and equality attract equal respect and none is superior to the other.[24]

It is important to recognize that hate speech prohibits the advocacy of hatred based on race, ethnicity, gender or religion and that constitutes incitement to cause *harm*. Expression that simply offends segments of the community does not amount to hate speech as defined in section 16(2)(c). A distinction must be drawn between an expression that offends and expression that harms or is likely to cause harm in the manner prescribed above. [25]

---

20 S 130.

21 S 319(1) of the Canadian Criminal Code.

22 In the classic case of *Chaplinsy v Hampshire* (315 U.S. 568 (1942)) fighting words were defined as those that neither contributed to the expression of ideas nor possessed any social value in the search for truth and that incited an immediate, violent response.

23 2003 11 BCLR 1283 (SAHRC).

24 1288D.

25 See also *Islamic Unity Convention v Independent Broadcasting Authority* 2002 5 BCLR 433 (CC) as quoted on 1292J of *Freedom Front*.

After discussing the legal position the commission came to the conclusion that:

> "The issue in respect of section 16(2)(c) of the Constitution is whether a reasonable person, assessing the advocacy of hatred on the stipulated grounds within its context and having regard to its impact and consequences would objectively conclude that there is a real likelihood that the expression causes harm. The closer the proximity or causal link between the advocacy of hatred on the stipulated grounds and the harm the more likely it is that the expression would be deemed to be hate speech. The more tenuous the proximity or causal link, the less likely it is that the expression would be deemed to be hate speech. There must be real likelihood that the expression causes harm before it can be deemed to be hate speech."[26]

The commission came to the conclusion that the slogan "Kill the Farmer, kill the Boer", in the context in which it was chanted,[27] "would harm the sense of well being, contribute directly to a feeling of marginalization, and adversely affect the dignity of the Afrikaners".[28] As such it was held to fall within the meaning of section 16(2)(c).[29]

The finding that a statement falls within the ambit of 16(2) does not mean that it is automatically prohibited.

Govender remarked in *Freedom Front* the following in this regard:

> "expressions falling within the categories listed in sec 16(2) cease to be protected. The implication of it not being protected is that once expression is deemed to be, for instance, hate speech, it may be regulated or totally proscribed by the State, provided it is rational to do so.
>
> The State, in this instance, does not have to justify the limitation in terms of the limitation clause as such regulation or proscription would not amount to an infringement of a constitutionally protected right."[30]

Once it is found that a statement is not protected by section 16(2), the next part of the enquiry is whether it is prohibited by the state. It is trite that the fact that a statement falling within the ambit of section 16(2), is of no consequence if it is not prohibited by legislation or the common law. Statements such as the above may, however, fall foul of the Promotion of Equality and Prevention of Unfair Discrimination Act 4 of 2000. According to its preamble, the Act was promulgated "to give effect to section 9 read with item 23(1) of Schedule 6 to the Constitution of the Republic of South Africa 1996, so as to prevent and prohibit unfair discrimination and harassment; to promote equality and eliminate unfair discrimination; to prevent and prohibit hate speech; and to provide for matters connected therewith". Section 10 of the Act prohibits hate speech and reads as follows:

> "10(1) Subject to the provision of sec 12, no person may publish, propagate, advocate or communicate words based on one or more of the

---

26 1298A.

27 It was specifically stated that "the context and content of the expression will have to be assessed on a case-by-case basis to determine whether it causes or is likely to cause harm as opposed to offence" (1296C).

28 1299D.

29 1299C.

30 1289C.

> prohibited grounds, against any person that could reasonably be construed to demonstrate a clear intention to –
> (a) be hurtful;
> (b) be harmful or to incite harm;
> (c) promote or propagate hatred."[31]

The proviso in section 12 provides that *bona fide* engagement in artistic creativity, academic and scientific inquiry, fair and accurate reporting in the public interest or publication of any information, advertisement or notice in accordance with section 16 of the Constitution is not precluded by this section.

Since the Equality Act's definition of hate speech is broad,[32] it is submitted that the words used in the statement under discussion clearly advocate hate based on race or ethnicity and constitute incitement to cause harm and therefore not only fall within the category of hate speech that is unprotected in terms of section 16(2), but also fulfil the requirements of hate speech that is prohibited by section 10 of the Promotion of Equality and Prevention of Unfair Discrimination Act. The words further clearly fall outside the ambit of section 12. One should therefore be able to take action against anyone who uses these words if one attaches the normal grammatical meaning thereto.

The situation becomes slightly more problematic if one is confronted with the fact that the statement is made on the Internet and, more specifically, on a social network such as Facebook. The first stumbling block in any action against perpetrators who make statements on a website external to South Africa, is the possible lack of jurisdiction. "The Internet has no territorial boundaries. To paraphrase Gertrude Stein, as far as the Internet is concerned, not only is there perhaps 'no there there' the 'there' is *everywhere* where there is Internet access."[33] Any action on the Internet reaches the whole of the Internet. This is particularly true of social network sites such as Facebook.

The question then is whether publication of a statement on Facebook can suffice to vest jurisdiction in South Africa. In terms of section 19 of the Supreme Court Act 59 of 1959, a court has jurisdiction over all persons residing or being in, and in relation to all causes arising and all offences triable within its area of jurisdiction. It is common cause that Facebook is based in Palo Alto, California in The United States. On the face of it therefore, publication on Facebook is external to South Africa and one could argue that a South African court will not be vested with jurisdiction to hear matters pertaining to statements on Facebook. There is, however, scope for another approach. In the only South African case dealing with the question

---

[31] As Currie and De Waal *The Bill of Rights Handbook* (2005) 378, points out, this "wordy" (378) section is not as clear as s 16(2) of the Constitution and it widens the scope of the prohibition far beyond the three clear grounds mentioned in s 16(2) of the Constitution to include discriminatory speech. It follows that speech falls foul of s 16(2)(c) of the Constitution will fall within the ambit of s 10.

[32] See, for instance, *Sonke Gender Justice Network v Malema* 2010 7 BCLR 729 733 par 13.

[33] Per Gertner J in *Digital Equipment Corp. v. Altavista Technology, Inc.*, 960 F.Supp 456, 462 (D. Mass 1997). See also *Blumenthal and Blumenthal v Drudge and American Online Inc.* 992 F Supp 44 (D. Col 1998).

of jurisdiction in a matter where words were published in cyberspace, the court assumed jurisdiction. In *Tsichlas and Another v Touch Line Media (Pty) Ltd*,[34] Kunay AJ decided, with reference to Burchell,[35] that publication for purposes of the law of defamation was within the area of jurisdiction of the court because the site on which the alleged defamatory statement was made, was accessed and was accessible in the jurisdictional area of the court.[36] The court then went on to say:

> "I do not propose to deal with the various complications which may arise from this finding. In effect, my conclusion would mean that, whenever anybody, anywhere in the world, accesses this website and reads and understands the words which are complained of in this matter, there will have been publication to that user at the particular place where the user has accessed the website. Bearing in mind that we are dealing with the Internet and electronic communications, that national or geographic boundaries would not apply and that distances are irrelevant, the implications of this conclusion are enormous."

In terms of *Tsichlas* the speech *in casu* was published within the area of jurisdiction of a South African court because it was accessable in South Africa. It was further directed at a group within South Africa. A South African court will therefore be vested with jurisdiction bearing in mind that the cause of action (publication) arose in South Africa and the defendant is present in the jurisdictional area of the court. Section 10(2) of the Promotion of Equality and Prevention of Unfair Discrimination Act is thus applicable and the remedies mentioned in this section are available to the aggrieved party against the offender.

The question whether an aggrieved party will have a remedy against the service provider (Facebook in this case) will depend upon, *inter alia*, the question how hate speech on the Internet is treated internationally. This matter will be addressed below.

5 INTERNATIONAL RESPONSES TO HATE SPEECH IN CYBERSPACE

Although most states have legislation governing hate speech, dissemination of hate speech over the Internet is rife.[37]

Some of the problems with regulation of the Internet is that it is borderless, it provides for anonymity and it is multijurisdictional.[38] The effect of the above is that a victim often finds it difficult, if not impossible, to act against a wrongdoer, even if his identity is known, because of difficulties with jurisdiction. Where states legislate against Internet abuses, perpetrator websites often simply move their websites to jurisdictions where there is less

---

34 2004 2 SA 112 (W).

35 Burchell *The Law of Defamation in South Africa* (1985) 79.

36 120.

37 For a discussion of this phenomenon see *Legal Instruments for Combating Racism on the Internet*, a report prepared and updated by the Swiss Institute of Comparative Law (Lausanne). Council of Europe Publication Strasbourg 2009.

38 Blarcum 2005 62 *Washington & Lee LR* 783.

or even no regulation. The result is the existence of so-called "safe havens" where Internet service providers are more or less untouchable.[39]

There are two ways in which states can address the "safe haven" problem. Firstly, Internet traffic from an unsavoury source can be blocked. This can be done either by blocking all services by a specific service provider or by blocking specific content.[40] Apart from the fact that it is technically difficult to block specific content from a service provider effectively, the common opinion is that the blocking of services or service providers goes against the idea of the free flow of information and freedom of expression. Such measures are regarded as extreme and not easily resorted to.

The second possibility is to deal with specific problems such as hate speech, pornography and child trafficking on an international basis through international law. In 2001 for instance, the World Conference against Racism, held in Durban, resolved that the International Convention on the Elimination of All Forms of Racial Discrimination (mentioned in par 3 above) should apply to the Internet.[41]

In the Council of Europe, criminalization of certain actions in cyberspace was first brought about by the Convention on Cybercrime, in 2001.[42] According to Akdeniz[43] the fact that not only the Council of Europe, but also the USA, Canada, South Africa and Japan, were involved in preparing the text of the Convention, "reflects both the international nature of the Internet and the fact that the economic and technical know-how gives them a remarkable capacity to shape how the Internet functions". The Convention focuses primarily on copyright issues, computer fraud, child pornography and violations of network security[44] and was signed (but not ratified) by South Africa in 2001.[45] The regulation of racist and xenophobic acts through the Internet was discussed, but not implemented by the Convention, because it was clear that the USA would be reluctant to sign the Convention if these matters were included in the Convention.[46]

In 2003, however, the Council adopted the Additional Protocol to the Convention on Cybercrime concerning the criminalization of acts of a racist

---

39 Safe havens in this sense must be distinguished from safe harbours created by legislation such as the Community Decency Act in the United States. Safe havens normally exist where there is very little or no regulation of the Internet.

40 *Eg*, Spain has recently blocked sites which do not comply with its national laws. See Blarcum 2005 62 *Washington & Lee LR* 784.

41 Rorive 2009 17 *Cardozo Journal of Comparative and International Law* 420.

42 European Treaty Series, No 185 http://conventions.coe.int/Treaty/en/Treaties/Html/185.htm (accessed 2010-07-20).

43 Akdeniz *Legal Instruments for Combating Racism on the Internet* (2009) Council of Europe Publishing 132. See also Akdeniz 2007 56 *University of New Brunswick LJ* 103.

44 Blarcum 2005 62 *Washington & Lee LR* 784.

45 Roos 449.

46 *Ibid*.

and xenophobic nature committed through computer systems.[47] As the name indicates, the purpose of the Convention was to make provision for the criminalization of acts of a racial or xenophobic nature committed through computer systems. The Convention defines racist and xenophobic material as "any written material, any image or any other representation of ideas or theories, which advocates, promotes or incites hatred, discrimination or violence, against any individual or group of individuals, based on race, colour, descent or national or ethnic origin, as well as religion when used as a pretext for any of these factors".[48] This Convention was signed (but not ratified) by South Africa in 2008. Article 3 of the Convention requires parties to criminalize the intentional distribution or making available of racist and xenophobic material to the public through a computer system.[49] It is clear from the definition of "racist and xenophobic material" that the material must promote or incite hate, discrimination or violence based on race, colour, decent or national or ethnic origin or religion before its distribution must be criminalized. Article 4 requires states to criminalize conduct which threatens, through a computer system, persons or groups belonging to a specific group distinguished by race, colour, descent or national or ethnic origin, as well as religion, with serious criminal conduct.[50] Article 5 requires criminalization of conduct which, through a computer system, publicly insults persons for the reason that they belong to one of the groups.[51] The protocol further makes provision for the extradition of persons between convention states by making the general extradition provisions of the Convention on Cybercrime applicable to the protocol.

Once states have aligned their legislation with the Convention on Cybercrime and more specifically with the Additional Protocol, it will be possible to act much more effectively against perpetrators of hate speech in Europe and other states who have legislated in terms of these instruments.

The question whether Internet service providers can incur liability under the prohibitions in terms of the protocol depends on their intent. Blarcum

---

47 Additional Protocol to the Convention on Cybercrime, Concerning the Criminalisation of Acts of a Racial or Xenophobic Nature Committed through Computer Systems. See http://conventions.coe.int/Treaty/en/Treaties/Html/189.htm (accessed 2010-07-21).

48 Article 2.

49 Article 3: Each Party shall adopt such legislative and other measures as may be necessary to establish as criminal offences under its domestic law, when committed intentionally and without right, the following conduct: distributing, or otherwise making available, racist and xenophobic material to the public through a computer system.

50 Article 4: Each Party shall adopt such legislative and other measures as may be necessary to establish as criminal offences under its domestic law, when committed intentionally and without right, the following conduct: threatening, through a computer system, with the commission of a serious criminal offence as defined under its domestic law, (i) persons for the reason that they belong to a group, distinguished by race, colour, descent or national or ethnic origin, as well as religion, if used as a pretext for any of these factors, or (ii) a group of persons which is distinguished by any of these characteristics.

51 Article 5: "Each Party shall adopt such legislative and other measures as may be necessary to establish as criminal offences under its domestic law, when committed intentionally and without right, the following conduct: insulting publicly, through a computer system, (i) persons for the reason that they belong to a group distinguished by race, colour, descent or national or ethnic origin, as well as religion, if used as a pretext for any of these factors; or (ii) a group of persons which is distinguished by any of these characteristics."

argues[52] that they can be liable under the doctrine of "permissive intent" if they are made aware of material on their server and they do not take action to remove it.[53]

Another international instrument of importance in the fight against cyber hate is the European Directive on E-commerce which came into operation in 2002. The Directive limits civil liability of ISPs for content on their servers, but the immunity is not a complete immunity as the immunity afforded by the Communications Decency Act in the United States.[54] The provider of hosting services can be held liable for damage if, after obtaining actual knowledge of the infringing material, it does not act expeditiously to remove the material.[55] The actual knowledge requirement can also be satisfied if an aggrieved party gives notice to the service provider of infringing material on the site. The Directive thus makes provision for a notice and take-down procedure. The effect of these procedures is that, once a service provider received a request to remove infringing material from its site (a take-down notice), and it does not expeditiously act upon it, it can be regarded as having actual knowledge of the infringing material and can thus be held liable. Almost all states in Europe have enacted legislation in accordance with the Directive on E-commerce. According to Rorive,[56] the provisions of legislation based on the Directive on E-commerce have been used by states in Europe to encourage service providers in the United States with assets in Europe to act on take-down notices.[57]

From the above it can be concluded that the international community, especially Europe, Canada, Japan, but also South Africa, has been involved on an international level to combat the dissemination of hate speech through the Internet. The question remains whether these instruments will result in effective enforcement at state level. It is submitted that it will be possible to act against individuals who make themselves guilty of spreading hate speech in most states which have enacted legislation in terms of the Convention on Cybercrime and the Additional Protocol to the Convention on Cybercrime concerning the criminalization of acts of a racist and xenophobic nature committed through computer systems.

In South Africa the Department of Justice is in the process of preparing a bill that will make South African law tougher on hate crimes. It will create offences relating to, amongst others, hate speech.[58]

---

52 Blarcum 2005 62 *Washington & Lee LR* 795.

53 This is also in line with s 14 of the European Union's Directive on Electronic Commerce which came into force in 2002. The directive limits liability of service providers which store information unless the provider has knowledge of an illegal activity. Directive 32000L0031 http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000L0031:EN:HTML. See also Rorive 2009 17 *Cardozo Journal of Comparative and International Law* 422.

54 Communications Decency Act of 1996. See discussion in par 6 1 below.

55 Article 14.

56 2009 17 *Cardozo Journal of Comparative and International Law* 423.

57 The argument goes that, if illegal content is published in Europe in a state where an American service provider has business interests, such provider will be granted immunity when it complies with a warning that it is hosting illegal content (Rorive 2009 17 *Cardozo Journal of Comparative and International Law* 423).

58 News 24 of 19 July 2010.

6 NETWORK NEUTRALITY AND INTERMEDIARY LIABILITY OF INTERNET SERVICE PROVIDERS

The question whether a provider of third party content, such as Facebook, MySpace or Twitter could be held liable for hate speech published on their website or be forced to take down such speech, requires an investigation into service provider liability, sometimes referred to as intermediary liability,[59] for illegal content. As will be shown below, during recent years many states enacted legislation to protect Internet service providers against liability for content put on their servers by third parties.

6 1 United States[60]

In the United States hate speech is regarded by some as the price society has to pay to safeguard freedom of expression.[61] In terms of the First Amendment to the United States constitution, "Congress shall make no law … abridging the freedom of speech or the press …"[62]

The prevalence of the First Amendment finds expression in the Communications Decency Act (CDA) of 1996. The CDA creates so-called safe harbours where Internet service providers are safeguarded from liability for *inter alia* content posted by third parties on their servers. Section 230 of the CDA, exempts ISPs and other users of interactive computer services from being held responsible for excessively violent, defamatory or harassing material, posted by third party users on their servers. In terms of section 230 of the Act, a provider of an interactive computer service shall not be regarded as a publisher or speaker of such content. The protection afforded by section 230, as will be shown below, can be regarded as the most comprehensive in the world.

In *Jane Doe v MySpace Inc*,[63] Clement (Circuit Judge) put it as follows:

> "In October 1998, Congress recognised the rapid development of the Internet and the benefits generated by web-based service providers to the public. Sec 47 U.S.C. par 230(a) (acknowledging that 'interactive computer services offer a forum for a true diversity of political discourse, unique opportunities for cultural development, and myriad avenues for intellectual activity' and have 'flourished ... with a minimum of government regulation'). In the light of its findings, Congress enacted the CDA for several policy reasons, including 'to remove disincentives for the development and utilization of blocking and filtering technologies that empower parents to restrict their children's access

---

59 Azriel "Social Networking as Communications Weapon to Harm Victims: Facebook, Myspace and Twitter Demonstrate a Need to Amend Section 230 of the Communications Decency Act" 2009 26 *The John Marshall Journal of Computer & Information Law* 415.

60 The law relating to service provider liability for defamation in the United States before the Community Decency Act is well documented (see amongst others, Carter, Dee and Zuckman *Mass Communication Law in a nutshell* (2007) 57; and Powers *The Internet Legal Guide* (2002) 50. It is not the intention to discuss this development here.

61 Bakircioglu "Freedom of Expression and Hate Speech" 2008-2009 16 *Tulsa Journal of Comparative and International Law* 1.

62 US Constitution Amendment I.

63 474 F.Supp.2d 843 (W.D. Tex. 2007)); 2008 WL 2068064 (5th Cir. May 10, 2008).

> to objectionable or inappropriate online material'. Id. par 230(b)(4). To achieve that policy goal, Congress provided broad immunity under the CDA to web-based service providers for all claims stemming from their publication of information created by third parties, referred to as 'Good Samaritan' provision. Id. par 230(c)(1) ('No provider or user of an interactive computer service shall be treated as a publisher or speaker of any information provided by another content provider.') Indeed, 'no cause of action may be brought and no liability may be enclosed under any State or local law that is inconsistent with this section'. Id par 230(e)(3)."

The CDA further specifically provides that a service provider will not be liable on the ground that it voluntarily and in good faith took steps to restrict access to or availability of objectionable material.[64] It is therefore not possible to argue that, because the service provider took steps to control content on the site, it is a provider of content on that site. Many service providers make use of so-called "Good Samaritan"[65] clauses. These are clauses by which service providers reserve the right to remove or restrict access to material which they regard as objectionable. The CDA further does not contain provisions for take down notices.

In *Doe v MySpace,* it was held that the CDA protects web-based service providers from liability even after the provider is notified of the objectionable content on its site and it refuses to remove it.[66] The court, quoted *Zeran*,[67] where the Fourth Circuit wrote:

> "If computer service providers were subject to distributor liability, they would face potential liability each time they receive notice of a potentially defamatory statement – from any party, concerning any message ... Because service providers would be subject to liability only for the publication of information and not for its removal, they would have a natural incentive simply to remove messages on notification, whether the contents were defamatory or not. Thus, like strict liability, liability upon notice has a chilling effect on the freedom of Internet speech ... Because the probable effects of distributor liability on the vigor of Internet speech and on service provider self-regulation are directly contrary to \$230's statutory purposes, we will not assume that Congress intended to leave liability upon notice intact."

Social network sites such as Facebook and Twitter are thus protected by their respective "Good Samaritan" clauses.[68]

---

64 S 230(c)(2) deals with civil liability and provides:
"No provider or user of an interactive computer service shall be held liable on account of –
(A) any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or
(B) any action taken to enable or make available to information content providers or others the technical means to restrict access to the material described in paragraph."

65 A "Good Samaritan" clause is a provision that protects a party who undertakes to protect another without being obliged to do so.

66 Report of the 5th Circuit Court of Appeals 8.

67 *Zeran v America Online Inc* 129 F.3d 327 (4th Cir 1997) 333.

68 See Azriel 2009 26 *The John Marshall Journal of Computer & Information Law* 421. Par 5 of Facebook's terms and conditions deals with "Other peoples' rights" and provides *inter alia* the following:

In comparison with Europe and South Africa, the United States undoubtedly affords the most protection to Internet service providers. As pointed out previously, this is a typical result of the prevalence of the First Amendment in the United States.

6 2 The Council of Europe and the United Kingdom

As mentioned above, the Council of Europe adopted the Convention on Cybercrime in 2001,[69] and the Additional Protocol to the Convention on Cybercrime concerning the criminalization of acts of racist or xenophobic nature committed through computer systems came into operation in 2006 and was signed by 30 states. Further, in 2001, the World Conference against Racism held in Durban, made it clear that the UN Convention on The Elimination of All Forms of Racism should also apply to the Internet.

In the United Kingdom, authorities have indicated that ISPs can be prosecuted in terms of the Public Order Act discussed above where they are aware of hate speech on their networks and fail to remove it.[70] It is therefore possible for an Internet service provider to be liable as a publisher or distributor in terms of the Public Order Act of 1986. There have, however, been no prosecutions in the United Kingdom for hate speech on the Internet.[71] An Internet service provider will have a defence in relation to the offences created in terms of the Public Order Act if it can prove that it did not have the intention, or did not reasonably suspect the words published on its website, were threatening, abusive or insulting.[72]

In 2002 the United Kingdom adopted the Electronic Communications (European Community) Regulations in order to implement the EU Directive on E-Commerce.[73] The regulations provide that an Internet Service Provider will not be held liable if it is only caching information, acting as a mere conduit or merely acting as a host for third party content. Such Internet service provider's liability is limited in terms of the EC Regulations[74] should they unknowingly transmit or store third party unlawful content. Although these defences deal with unlawful content in general, it is submitted that it will cover hate speech. The regulations provide that the protection is available as long as the service provider:

"1. You will not post content or take any action on Facebook that infringes or violates someone else's rights or otherwise violates the law.
2. We can remove any content or information you post on Facebook if we believe that it violates this Statement."

69 Signed by 38 members including Canada, Japan and South Africa. See Rorive 2009 17 *Cardozo Journal of Comparative and International Law* 421.

70 *Legal Instruments for Combating Racism* 96.

71 Nel "Freedom of Expression and the Internet" in Buys, Cronje (eds) *Cyberlaw@*SA (2004) 223; and *Legal Instruments for Combating Racism* 96.

72 S18-23 and s 29B-G of the Public Order Act of 1986.

73 Lloyd *Cyber Law in the United Kingdom* (2010) 166.

74 *Legal Instruments for Combating Racism* 97. S 21 of SI 2002/2013 expands the s 17-19 categories of mere conduits, caching and hosting to criminal proceedings as well. See also Lloyd (2010) 166.

(i) does not have actual knowledge of unlawful activity or information and, where a claim for damages is made, is not aware of facts of circumstances from which it would have been apparent to the service provider that the activity or information was unlawful; or

(ii) upon obtaining such knowledge or awareness, acts expeditiously to remove or disable access to the information …"[75]

If an Internet Service Provider therefore knowingly transmits hate speech or fails to react to a take-down notification, it loses the protection of the EC Regulations and commits an offence in terms of the Public Order Act.

6 3 Republic of South Africa

In South Africa, Chapter XI of the Electronic Communications and Transactions Act 25 of 2002 deals with "limitation of liability of service providers". The term "service provider" carries a wide meaning and means according to the Act "any person providing information system services."[76] Section 71 provides that before a service provider can make use of this limited protection it must belong to a so-called industry representative body recognized by the Minister of Communications and the Internet service provider must have adopted and implemented the code of conduct of such body.[77] It therefore clearly deals with South African Internet service providers only.

Internet Service Providers who have complied with the requirements mentioned above may make use of the "safe-harbour" provisions of the Act. They will not be liable for unlawful content posted by third parties (including hate speech), if they perform certain functions in a particular manner in relation to material that is placed on their websites.[78] The "safe-harbour provisions" covers only situations where third parties posted content on a website and where the Internet service provider is only a passive provider.[79] An Internet service provider acting as a host for information must further have designated an agent to receive notifications of infringement.[80] The immunity granted in terms of Chapter 11 of the Act is further conditional in that the service provider must, upon receipt of a take-down notification, remove or prevent access to the data alleged to be unlawful.[81] A further condition for immunity is that the service provider must not have actual knowledge of the infringing nature of the data or knowledge of facts or

---

75 Article 15 of the Directive.

76 S 70. "Information system services" is defined in s 1 of the Act as including "the provision of connections, the operation of facilities for information systems, the provision of access to information systems, the transmission or routing of data messages between or among points specified by a user and the processing and storing of data, at the individual request of the recipient of the service".

77 S 72.

78 S 73-76.

79 The act distinguishes between service providers which are mere conduits (s 73), those that cache information during transmission (s 74) and those that act as hosts (s 75).

80 S 75(2).

81 S 75(3).

circumstances from which the infringing nature of the data is apparent.[82] The Act further makes it clear that there is no general obligation on an Internet service provider to monitor data which it transmits or stores.[83]

It is noteworthy that the "safe harbour provisions" in the Act are similar to those in the E-Commerce Directive. This brings the South African law in line with that of the EU on this issue.[84]

7 TRANS-BORDER REGULATION

The Convention on Cybercrime, the Additional Protocol and also the E-Commerce directive require that legislation in the EU states be aligned to reflect the spirit of the directives. An example of such an alignment is the English Electronic Commerce (Directive) Regulations discussed above. One can therefore assume that EU states can act in concert and that, should hate speech appear, even on Internet service providers of social network sites in the EU, Internet service providers may incur liability if they are made aware of infringing content and do not act to remove it. The question is whether these measures can be used to curb hate speech in non-aligned states, especially in the United States. Can the European Directive set up a notice and take-down procedure that can reach American hosting services?[85]

According to Rorive the combination of the "Good Samaritan clause" of the CDA and the European Directive makes it possible to reach American Internet companies with financial interests in Europe, despite the First Amendment.[86] He mentions two instances where such a tool has been successfully used. In 2001 and 2002 Germany took action against eBay, a large shopping site. The German Agency for the Protection of the Constitution warned eBay against the sale of Nazi songs, books, clothes and other paraphernalia. In both cases, eBay reacted by disabling access to the items and declared formally that it "will no longer host the sale of memorabilia of the Nazi period or anything related to fanatical groups".[87] The second instance is one where a French court ordered, under the threat of a fine of 500 Euros per day, three US companies to cut access to a revisionist website called AAARGH. The judge also ordered approximately 10 French hosting providers to cut access to the site.[88]

As can be seen from the above examples, the regulation of hate speech in Europe can and does have an effect, be it indirectly, on the safe harbours in the United States.

Will it be possible to act effectively in a South African court against a social network site such as Facebook? If one assumes that a South African

---

82 S 75(1)a and b.
83 S 78.
84 For a detailed discussion see Roos 418-434.
85 Rorive 2009 17 *Cardozo Journal of Comparative and International Law* 423.
86 *Ibid*.
87 Rorive 2009 17 *Cardozo Journal of Comparative and International Law* 424.
88 Rorive 2009 17 *Cardozo Journal of Comparative and International Law* 425.

court will have jurisdiction because the statement complained of was made by a South African resident while in the jurisdictional area of the court and it was "published" in South Africa where it was accessible in terms of *Tsichlas v Touch Line Media (Pty) Ltd (supra),* it may be possible to convince a court in South Africa to make an order against Facebook.

The question whether a social network site has a sufficient presence within the area of jurisdiction of the court for it to assume jurisdiction, depends on the facts of the particular situation. If the defendant has a physical presence in the area of the court in the form of an office or other assets, the court may assume jurisdiction to hear the matter. In *casu*, Facebook, unlike for instance Google, does not have a physical presence in South Africa and it is therefore doubtful whether a court will allow actions against Facebook. This is the first hurdle.

The second problem deals with the question whether a judgment obtained in South Africa, will be enforceable against Facebook. This brings one within the realm of private international law. Although it is common cause that the merits of a judgment which one attempts to enforce is normally not relevant as a factor to decide whether the judgment is enforceable, US courts have consistently refused to enforce English judgments pertaining to defamation against US citizens. The attitude of the US courts is simply that such judgments go against the First Amendment and are therefore against the public policy of the United States law.

In the light of the above, even if it would be possible to obtain a judgment in a South African court against Facebook based on hate speech, it will not be enforceable in the United States. Although a claim based on hate speech will not be subject to a choice of jurisdiction or choice of law clause if one is not subject to the terms and conditions of use of Facebook, it is nevertheless interesting to note Facebook's choice of jurisdiction clause, which provides that one subjects oneself to the state or federal court of the county of Santa Clara and to the laws of the state of California where the First Amendment rules.[89]

8 CONCLUSION

Great strides have been made over the last decade or two to curb hate, racist and xenophobic speech in Europe and the rest of the world. The regulation of hate speech is based on International instruments such as the UN *Convention on the Eradication of All Forms of Racism*, the EU's Convention on Cybercrime and more specifically the Additional Protocol to the Convention on Cybercrime, concerning the criminalization of acts of a

[89] Par 15 of Facebook's terms of service deals with "disputes". Par 15.1 provides: "You will resolve any claim, cause of action or dispute ('claim') you have with us arising out of or relating to this Statement or Facebook exclusively in a state or federal court located in Santa Clara County. The laws of the State of California will govern this Statement, as well as any claim that might arise between you and us, without regard to conflict of law provisions. You agree to submit to the personal jurisdiction of the courts located in Santa Clara County, California for the purpose of litigating all such claims."

racist or xenophobic nature committed through computer systems. These instruments resulted in legislative regulation of hate speech, not only in many states in Europe, but also in states such as South Africa and Canada. Hate speech on the Internet, and especially on social network sites, is relatively effectively addressed in the EU and many other Western states.

In the United States social network sites are protected by the safe-harbour provisions of the Community Decency Act. In states where social network sites are protected or not subject to hate speech regulation, one is dependent on self-regulation by the Internet Industry. A good example of self-regulation can be found in the Terms of Service of Facebook. Under the heading: "Statement of rights and responsibilities", paragraph 5 deals with "Protecting other peoples' rights" and provides that it can remove harmful content from its site. [90] The author is aware of many instances where Facebook has removed content where it was brought to its attention that the material was harmful.

In conclusion, on the whole one can agree with Akdeniz:

> "The myth of the Internet as a lawless, godless zone should be dismissed at the outset. This myth of a legal vacuum – which is encouraged by certain alarmist politicians, amplified by the press and exacerbated by ill-considered declarations of independence by "surfers" eager for absolute freedom – does not stand up to scrutiny. Like any other means of communication, the Internet has to function within the law."[91]

[90] Par 5; and see fn 68 above.

[91] *Legal Instruments* 42.