Assessing Trust: Contextual Accountability

Matthew Rowe and Jonathan Butters

OAK Group, Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield, United Kingdom

{m.rowe,j.butters}@dcs.shef.ac.uk

Abstract. The need to assess the trustworthiness of a piece of information has become a prevalent issue due to web vandalism and the openness of platforms such as wikis. In this paper we introduce the notion of contextual accountability: Holding the author of a given statement accountable by deriving a path from the statement to the author's credentials. We present our own trust vector model as a combination of the assessment of these credentials covering two aspects; knowledgeability and social acceptance. This model is put to the test using two datasets and the results are presented and discussed.

1 Introduction

With the abundance of user generated content on the web comes the need to filter or weight the validity of statements and factoids. This ability is innate in humans but is completely lacking in agents and software that blindly rely on linked data and triples as hard fact. There are many types of users that contribute information to online sources, and these can generally fall into positive contributors and negative contributors. Negative contributors can be subdivided into three types: Troll - a knowledgeable user who posts negatively biased posts to create angst amongst the community, Spammer - a non-knowledgeable user who floods the community with unrelated posts usually offering a product or service, and Truth Engineer - a highly knowledgeable user who generally contributes to a very narrow domain, often trying to manipulate the community's point of view to better align with some ideal. It is therefore obvious that different levels of trust should be given to different statements made by authors based on the their areas of expertise and their posting habits within the sub ject area. In this paper we propose a methodology for assessing the level of trust given to statements made in a particular context by estimating the level of knowledge the author has in this field and how respected this author is by community members.

Wikipedia is a prime target for Trolls and Truth Engineers. The Wikipedia article regarding Cypriot football team AC Omonia included bogus information regarding a fan club lovingly called 'The Zany Ones', who wore 'shoes on their heads', had a song 'about potato' and 'kept their season tickets in the oven for safekeeping'. Attention was brought to these edits when they were inadvertently

taken as fact and printed in a national UK tabloid on the 18th September 2008[1]. The reliance on knowledge sources such as Wikipedia has also forced Wikipedia to rethink their edit policy for pages about living people and companies[2] following a spate of malicious edits on such pages. The proposed changes to the edit policy will see trusted community members vetting edits prior to their inclusion in the page.

In this paper we present our method for deriving trustworthiness of a person given a context, relying on holding a person accountable for their actions. Semantic Web technologies and formalisations allow graphs to be created containing links between individual semantic resources. Paths through such a graph space provide a technique whereby a given statement can be traced back to the statement author, thus holding that person accountable. We define accountability as the possibility to generate this path structure, which is only achievable through a semantic graph space using linked data across multiple data sources and knowledge bases. The bill of rights for web users[3] gives grounding to this theory by treating web users not simply as consumers but as citizens, making them responsible for their own actions. This is the stage that we believe the web has now reached, where data sources must be assessed for validity by assessing the statement authors and contributors.

Deriving a value of trustworthiness for a person given a context requires the assessment of known information that can support this person's claim as a trusted entity. In this paper we divide this information into two distinct aspects; knowledge and social. The former referring to formal information accredited to the individual such as publications or reports, essentially providing validation that this person's expertise is evident within the context in question. The latter refers to the community acceptance socially attributed to the person; in this case it can be message board posts, or blog posts critiqued by the community. We combine each aspect to provide a vector representation of the person's overall trustworthiness, depending on the position in the vector space, it is possible to derive a classification for the person in question and whether they should be trusted or not, given the context.

This paper is structured as follows: Section two presents an approach to model accountability paths through a semantic graph space from a given statement through to the credentials of the statement author. Section three describes the metrics used to derive trustworthiness from both the social and knowledge aspects of a statement author's credentials, and how these measures are combined to produce a vector representation. Section four describes the experiments conducted to demonstrate the effectiveness of our approach, and the discussion of the obtained results. Section five compares state of the art trust derivation techniques with our approach, and section six describes the conclusions drawn from this work and discusses planned future work.

---

[1] http://www.thespoiler.co.uk/index.php/2008/09/19/daily-mirror-football-journalist-merked-by-wikipedia

[2] http://technology.timesonline.co.uk/tol/news/tech_and_web/the_web/article5593986.ece

[3] http://opensocialweb.org/2007/09/05/bill-of-rights/

2 The Semantics of Accountability

In order to assess for trust we rely on the existence of linked data to derive a path through the web from the a given statement through to credentials attributed to the statement author. We define a statement as a loose abstract notion of a piece of knowledge added to a knowledge base (e.g. OWL Full is more expressive than OWL DL). In essence we wish to take a statement and derive a trust metric for this statement given two properties: The author, and the context. The credentials we define cover two different aspects: knowledge and social. The former deals with formal work attributed to the statement author, and the latter focuses on the community recognition attributed to the statement author. In each case, the goal is to assess each aspect with respect to the context of the statement in question.

2.1 Gathering a semantic representation

To contextualise this, imagine a person updates the Semantic Web Wikipedia page. In order to assess the trustworthiness of this update we assess the credentials associated with the author, holding them accountable for this update. Both the social and knowledge aspects are combined to derive a vector interpretation of the author's trustworthiness given the context of the statement, which in this case is about the Semantic Web. The derivation of this vector is explained in detail in the following section.

At a low level, holding a statement author accountable requires the derivation of a path through the web from the statement through to the credentials attached to the statement author. Linked data and Semantic Web formalisations enable such paths to be derived using the intrinsic graph structure of semantic formalisations as we will now explain. We begin by taking a statement to be analysed, given the nature of knowledge bases (e.g. Wikipedia, Freebase, etc), each statement can be related to an author representation within that platform, e.g. A profile page about the author. At this stage we assume that a semantic representation of this author exists as RDF according to the FOAF specification [2], however we must find this information. Therefore extracting this semantic representation is carried out using one of the three methods:

1. Query the Semantic Web for explicit representation: The author's handle or username is submitted to a Semantic Web entry point such as Watson[4] or Swoogle[5], and the relevant FOAF file is extracted.
2. Query the wider web for semantic representation: The author's handle or username is submitted to a web search engine such as Google[6]. The most relevant page is assessed for the existence of lightweight semantics such as Microformats[5] or RDFa [4], and explicitly linked FOAF files.

[4] http://watson.kmi.open.ac.uk/WatsonWUI
[5] http://swoogle.umbc.edu
[6] http://www.google.com

3. Implicit semantics within the existing profile page: Several platforms now support the use of RDFa and Microformats within XHTML. Given this use of implicit semantics, the profile page is parsed to derive credentials relating to the knowledge and/or social aspects, or linked data representations where such information can be found (ie. Hyperlink to a FOAF file).

2.2 Deriving Credentials

Given that we now have a semantic representation describing the statement author, we assume that this formalisation contains the credentials of each aspect or a linked data representation of such information. Credential information describing the knowledge aspect is formalised using the Bibtex specification[7], where each paper is an instance of `bibtex:Entry` or one of it's subclasses, and paper details are defined using `bibtex:hasTitle` for the title, `bibtex:hasKeywords` for keywords, and `bibtex:hasBookTitle` for the title of the publication. The context of the statement is analysed against these properties, given that we do not wish to assess all publications attributed to the author, simply the publications within the context of the statement, thus deriving a subset of the original publications set.

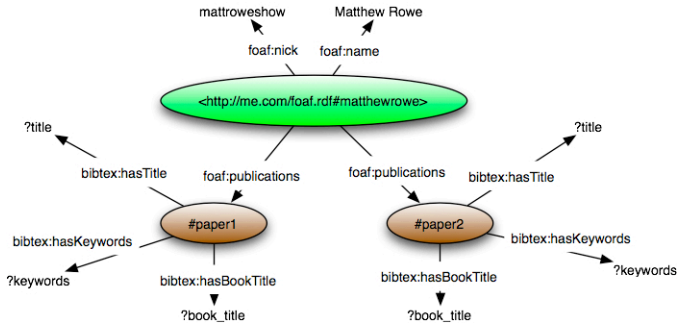

Fig. 1. Statement author linked to publications as instances of bibtex:Entry

Credential information related to the social aspect is derived using linked data within the FOAF file. The instance of `foaf:Person` describing the statement author is linked to an instance of `sioc:User` using the `sioc:account_of property`. As SIOC [1] is purposely designed to define online community structures, a weblog or forum is attributed to the statement author using the `sioc:owner_of` property property with the `sioc:User` as the domain and `sioc:Container` as the range. Any blog or forum posts are then made within this container, where each post is an instance of `sioc:Post`. Our intuition at this stage is that the blog attributed to the statement author will contain numerous posts about a variety

[7] http://zeitkunst.org/bibtex/0.1/bibtex.owl

of subjects, therefore in a similar manner to the knowledge aspect we derive a subset of the total set of blog posts containing instances of `sioc:Post` related to the statement context. This is fascilitated by the use of the sioc:topic property to denote tags assigned to the blog posts, we compare these properties against the context and filter out all instance of `sioc:Post` that do not match. Figure 2 demonstrates the RDF graph structure linking an instance of `foaf:Person` due several blog posts as instances of `sioc:Post`.

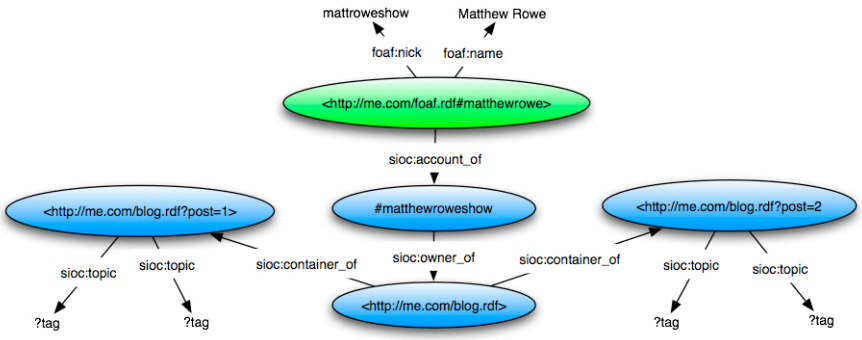

Fig. 2. Statement author linked to his/her blog posts as instances of sioc:Post

Once we have gathered the credential information spanning both the knowledge and social aspects, trust metrics must be derived using this information. We now discuss our proposed metrics with examaples demonstrating the derivation technique.

3 Deriving Contextual Trust

We define two orthogonal vectors; the social aspect, and the knowledge aspect, of which the vector addition results in the trust vector. The social aspect quantifies how respected the author's point of view is in a given domain by taking into consideration positive feedback and recent activity levels, similarity the knowledge aspect assesses the level of trust associated to scientific statements the author has made.

3.1 Social Aspect

We define the social trustworthyness of a given person within a certain knowledge topic as follows. Let p be the person in question, c is the topic of knowledge covered by the statement, P' is the set of posts (on a discussion board or blog) made by p about the topic c, F is the set of comments or feedback elements attributed to a given post. We further define the function $acceptance(f)$ as the classification of a feedback element describing acceptance in relation to the post

or not, this will return +1 if the feedback is positive, -1 if the feedback is negative and 0 if the feedback is neutral. $\left|P^{'}\right|$ denotes the number of posts made by that person about that topic, which we sqaure in order to increase the value of the posting more (different values were experimented with, squaring appeared to produce the optimum value). Therefore the social trustworthyness of a person given a knowledge context is as follows:

$$S(p,c) = \frac{\sum_{p' \in P'_{p,c}} (\sum_{f \in F_{p'}} acceptance(f)) \cdot (\left|P^{'}\right|)^2}{timeperiod}$$

3.2 Knowledge Aspect

We define the knowledge trustworthyness of a given person within a certain knowledge topic as follows. Let p be the person in question, c is the topic of knowledge covered by the statement. Q is the set of papers of which p is an author. With regards to the the knowledge topic c we wish to analyse the set of corresponding papers, we therefore define $R = Q \mid c$ as the set of papers within the knowlege topic, or context. We define a function $citations(r)$ that returns the number of positive citations the paper r has received: Each citation is assessed; positive citations receive +1, negative citations receive -1 and neutral citations receive 0. Using these defintions the knowledge trustworthyness of a person given a knowledge context is as follows:

$$K(p,c) = \frac{\sum_{r \in R} citations(r) \cdot \left(\sum_{y=1}^{x} \frac{|R_y|}{y}\right)^2}{timeperiod}$$

The formula takes in to account the recent activity of the person in question: The number of citations received in a certain year are divided by how many years ago the citations were made. This is denoted by $\frac{|R_y|}{y}$ where y ranges from 1 year ago to x number of years ago. Such a weighting is valid because within the Semantic Web domain, and indeed in other areas of web science, techniques and work moves extremely quickly. Therefore a more heavily cited recent piece of work should carry more weight than an older piece of work with the same number of citations.

3.3 Weighting Knowledge Aspects

As trust is a relative quantifier, the level of trustworthiness can only be calculated between the group of people who have made contributions to a particular resource, to this end, once the knowledge and and social scores for each person has been calculated for a particular context, each score must be normalised by the the magnitude of the maximum score in the sample set for each aspect. Trustworthiness is then represented by the resultant vector given by:

$$\overrightarrow{T} = S(p,c)\hat{S} + K(p,c)\hat{K}$$

where $\hat{S}$ and $\hat{K}$ are two orthogonal dimensions which form our vector space.

The resultant Trustworthiness vector $\overrightarrow{T}$ for each person will have a maximum potential magnitude of $\sqrt{2}$ and may lie in either of the four quadrants. By observing the location of each vector it is possible to assess the type of poster the user is. Individuals with higher positive social aspect scores tend to be more of a 'familiar name in the community' who's level of trust stems from the notion that they are observers (i.e. visible lurkers) within the context whose contribution to topic discussions are generally positive. Individuals with higher Knowledge aspect scores tend to be contributors to the context whose work tends towards to the state of the art.

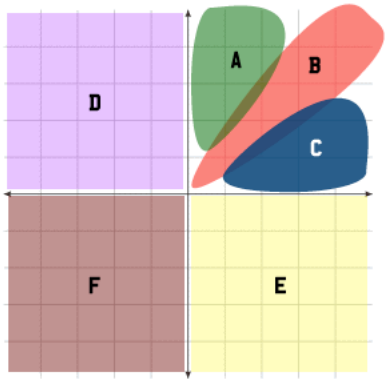

Fig. 3. Vector space representation of trustworthiness classifications. The x axis represents the knowledge score and the y axis represents the social score.

As it is possible for an individual to attain negative values for social and knowledge aspect vectors, it is possible for an individual to attain for instance, a negative value for the social aspect and a positive value for the knowledge aspect. This is indicative of a person who objects to points raised in conversations (such as challenging people's points of view with forum or blog posts), but who's work is original and is therefore cited.

Our intuition behind the design of the trust metrics and the resultant vector model is to allow classification of statement authors, and therefore reach a decision whether they should be trusted or not. If one considers figure 3, any person whose trust vector lies within the areas marked A, B or C can be considered a trustworthy person, with larger vector lengths representing a greater level of trust. However, should a person have a trust vector residing in the area marked D that classifies the person as having a good level of social trustworthiness, yet

their formal publications are not supported by the community (e.g. Einstein prior to his theory of relativity being proved by Ellington). Similarly, should the trust vector lie in the area marked E one could consider this person to be formally accepted as knowledgeable, yet socially their ideas are not accepted. (e.g. A truth engineer; CIA, geeks, nerds, etc). In either case, where the vector is in area D or E it is not certain whether this person should be trusted or not.

Finally, if the trust vector lies in the area marked F, it is fair to consider this person unstrustworthy due to their exclusion by the community socially and formally in their work. Our belief is that trolls and spammers would reside in this area given that their ideas are, in general, not supported by the community and their social trustworthiness is negative due to their motivations of using the web being for maliscious purposes.

4 Experiments

In order to assess the validity of our trust metrics we designed two experiments covering two separate domains of expertise; the Semantic Web and HCI (human computer interaction). In each domain we assumed that a wikipedia page had been edited and that statements within that page were linked to a set of people. In each case we derived the list of people based on conference proceedings and workshop organising committee lists to find those people who have both a blog and a list of publications. We then derived trust measures manually, checking positive and negative feedback and citations for the social and knowledge aspects respectively. The results were then collated and analysed.

4.1 Datasets

Two datasets were gathered for the experiments, where each dataset contained a list of people within a separate domain of expertise. For the Semantic Web dataset we obtained the list of potential Semantic Web statement authors from the linked data on the web workshop[8], social data on the web workshop[9] and european Semantic Web conference[10] participants lists. From these lists we then selected people with both papers and a blog, making sure that we chose a diverse range of experience and duration within the domain. A similar action was performed for the separate domain Human Computer Interaction (HCI), whereby authors of prominent papers were sampled.

4.2 Results

The results from the Semantic Web dataset demonstrate how for the people in the dataset, the majority carried a strong emphasis on blogging their work which

[8] http://events.linkeddata.org/ldow2009/
[9] http://sdow2008.semanticweb.org/
[10] http://www.eswc2009.org/

Name	Knowledge Aspect	Social Aspect	Relative Trustworthiness
Person A	0.085239676	0.251591844	0.265637592
Person B	0.393860438	0.683222914	0.788618789
Person C	1	0.007884645	1.000031083
Person D	0.120373117	1	1.007218788
Person E	0	0.008778798	0.008778798
Person F	0.12748067	0.363994004	0.385672084
Person G	0.121361371	0.111505714	0.164809304
Person H	0.090343433	0.070168176	0.114391909
Person I	0.15023741	0.026336395	0.152528309
Person J	0.27146477	0.011150571	0.271693682

Table 1. Trust statistics from the Semantic Web dataset

Name	Knowledge Aspect	Social Aspect	Relative Trustworthiness
Person K	1	0	1
Person L	0.22574209	0.007674504	0.225872507
Person M	0.76437217	0	0.76437217
Person N	0.079107224	1	1.003124096
Person O	0.219654267	0	0.219654267
Person P	0.120053261	0.034494332	0.124910545
Person Q	0.807120783	0	0.807120783
Person R	0.013191016	0.049538713	0.051264871

Table 2. Trust statistics from the human computer interaction dataset

Fig. 4. Vector space representation of Semantic Web dataset (left) and the HCI dataset (right). The x axis represents the knowledge score and the x axis represents the social score.

was in turn positively rated by the community of readers in the blogosphere. There are two evident extreme cases with respect to each aspect: Person D for the social aspect, and person C for the knowledge aspect. As the former blogs regularly it is no surprise that he sets the precedent for maximising trust value. So too in the case of person C, who has numerous publications all of which are heavily cited in the Semantic Web community. Figure 4 throws up some interesting questions though. Using our combination of each aspect measure into a single vector yields the derivation that person D is the most trustworthy within the Semantic Web dataset, following closely by person C.

Another interesting feature of these results is the correlation between the vectors of person A and person F. Although the former vector has less magnitude than the former, it falls along the same angle suggesting a correlation of trustworthiness with a similar spread between blog posts and publications.

Although the formula takes steps to mitigate the disparity between the level of trust extremely large numbers of citations and more average numbers of citations achieve, it can be seen that the Semantic Web results appear skewed on the knowledge aspect, this is due to the fact that one person was a prolific and heavily cited author of papers resulting in an overshadowing effect on the publishing counts of others. The HCI dataset results vary from the Semantic Web results due to the fact that many of the blogs that the HCI people maintain do not achieve levels of comments or feedback the Semantic Web blogs do, and therefore social trustworthiness in the HCI context is lower. However, as we define trustworthiness as the vector addition of two orthogonal aspects we maintain the ability to rank people according to their measured knowledgeability. The main reason for the higher number of people in the Semantic Web domain with social trust is most likely due to the Semantic Web experts' use of the blogosphere as a medium for publishing and sharing content as well as providing a quick way to expose ideas and discuss them. This increases other people's awareness of them as a positive influence, which will in turn increase their social trust score.

If we contrust figure 4 and figure 3, the positioning of the participants all lie in the upper right quadrant of the vector space. When analysed for negative feedback and citations, it became apparent that the majority were positive, and only a few negatives for each person. Our belief is that given a larger experiment incorporating a more controversial, and disputed domain of work, criticisms and negative comments would play a major role in deciphering trust with respect to a person and context.

5 State of the Art

In the Semantic Web community, work has been performed to investigate the network of trust necessary for Semantic Webs to function. Work by [3] proposes deriving trust values from an existing network too allow application and similar services that may wish to interface with the network to have access to an average trust value for the network. Unfortunately, altough each edge in the graph is denoted by a trust value, it is unclear how this value is derived. The notion

of trust is dealt with in work by [8] through the use of bilattice frameworks to compute trust associated with a given statement, similar to the goal of our work. [8] differs from this paper however, by utilising multiple trust levels depending on the data source hosting the statement. Semantic web services require trust to be estabished between the service requestor and provider, in such a problem trust information must be passed between the two parties. [6] describes such trust policies that when passed between the parties, must be met in order for trust to be achieved. Policies commonly contain a list of criteria in the form of logic statements. The majority of trust based research conducted within the Semantic Web utilises black box techniques to derive a trust measure for a given statement or person.

External to web semantics, communities such as security and privacy have provided approaches for modelling trust. With regards to the actual vector model demonstrated in this paper, work presented in [7] by Rat et al is comparable by providing a vector representation of trust associated with a given person within a given context as a combination of an experience aspect, knowledge aspect and cumulative effect aspect. Both our work and work by Ray et al utilises the context as an essential feature of trust derivation, however our work differs by computing relative trust scores among a group of statement authors. A very similar approach to our work is described by Kim et al in [9] by deriving the degree of trust for a given person from the affiliation and expertise the person has with the context. The model used by Kim et al is similar to our work, by creating a path from the statement author, which in the case of [9] is a reviewer, to their create content and the relevant replies.

6 Conclusions and Future Work

In this paper we have presented an approach to hold the author of a statement accountable, and provide a path of accountability from the statement to the author credentials through a semantic graph space. Our interpretation of trust as a vector allows the portable model of trust to be computed in a number of domains. As we have presented in this paper with the Semantic Web and HCI datasets. The exclusion of an evaluation of the derived trust metrics was largely due to the unavailability of sufficient evaluation participants at the time of conducting the experiments. It is vital for evaluation participants to have an understanding of the field in question – Semantic Web or HCI – in order to make an informed decision. We plan to investigate this further in the future.

Limitations meant that we were only able to manually examine a pre-chosen selection of people for the experiments resulting in our test individuals, scoring positive results. Future work will investigate experiments using additional datasets known to contain negative contributors, effectively producing a classification of the type of negative web user.

During the analysis of the results a third aspect of trust became more apparent. The notion that a small set of people may trust one another more because they are co-located. Although this aspect of trust is important in working rela-

tionships, where a higher level of direct accountability means that a person will generally assign a high degree of trust to those they work with, the requirement of a personal point of view makes the application of this trust metric to general statements impossible, unless of course the overall level of trust attained were only applicable to a single person's point of view, in this case, a third trust vector can simply be added. One addition to this separate aspect could include the notion of vouching. The chain of reliability in a trust network would denote that person A trusts person B, and therefore vouches for them. However, the level of trust assigned to person B could only be the maximum trust measure that person A currently has. Should that measure be low, then person B would only receive a weak trust measure from person A.

References

1. Uldis Bojars, Alexandre Passant, Richard Cyganiak, and John Breslin. Weaving sioc into the web of linked data. In *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008), Beijing, China*, Apr 2008.
2. Dan Brickley and Libby Miller. FOAF vocabulary specification. Technical report, FOAF project, May 2007.
3. Jennifer Golbeck, Bijan Parsia, and James Hendler. Trust networks on the semantic web. pages 238–249. 2003.
4. W3C Working Group. Rdfa primer: Bridging the human and data webs, October 2008.
5. R. Khare. Microformats: The next (small) thing on the semantic web? *Internet Computing, IEEE*, 10(1):68–75, 2006.
6. Daniel Olmedilla, Ruben Lara, Axel Polleres, and Holger Lausen. Trust negotiation for semantic web services. In *1st International Workshop on Semantic Web Services and Web Process Composition in conjunction with the 2004 IEEE International Conference on Web Services*, pages 81–95. Springer, 2004.
7. Indrajit Ray, Sudip Chakraborty, and Indrakshi Ray. Vtrust: A trust management system based on a vector model of trust. In *In Lecture Notes in Computer Science*, pages 91–105. Springer, 2005.
8. Simon Schenk. On the semantics of trust and caching in the semantic web. In *ISWC '08: Proceedings of the 7th International Conference on The Semantic Web*. Springer-Verlag, 2008.
9. J. Young Ae Kim Minh-Tam Le Lauw, H.W. Ee-Peng Lim Haifeng Liu Srivastava. Building a web of trust without explicit trust ratings. In *Proceedings from Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on*, pages 531–536, 2008.