ITU Workshop on
“Future Trust and Knowledge Infrastructure”, Phase 2
Geneva, Switzerland
1 July 2016

From weak online reputation metrics to standardized attack-resistant trust metrics

Dr. Jean-Marc Seigneur
President at Réputation SAS, Chief Reputation Officer at GLOBCOIN
Senior Lecturer and Research Manager at Medi@LAB, CUI ISS, G3S,
University of Geneva
Jean-Marc Seigneur@reputaction.com
Agenda

• Introduction
• Today’s Weak Online Reputation Metrics
• Computational Trust Engines
• Towards Standardized Attack-Resistant Trust Metrics
• Conclusion
• Q&A
Online reputation economy

- By 2026, thanks to online ratings
  - “a more successful hospitality and leisure sector has the potential to add an extra £2bn to the UK economy with the impact on the sector’s large supply chain contributing a further £1.2bn.” [Barclays, 2016]
Main online e-reputation ratings services for the general public

- Especially in the tourism industry
  - Around 60% of the hotel ratings by 2 providers only [TCI Research French, 2015]
    - Booking, whose ratings are verified because based after payment has been made, taking around 25% of the night cost
    - TripAdvisor, whose ratings are not verified
- Somehow: eBay, Yelp, Klout, TrustPilot, TrustYou, Facebook Reviews, Google Reviews...
Ratings for Google SEO

Val Thorens, France - Book and Save on Your Hotel Today

4.5 ★★★★★ rating for booking.com
Best Price Promise, No Booking Fee
Free Cancellation - Read Real Guest Reviews - We speak your language
Deals Special Offers No Booking Fees
Best Price Guarantee Book Now Pay Later

Altapura Hotel (Val Thorens) - 279 avis - TripAdvisor
★★★★★ Rating: 4.5 - 279 reviews - Price range: 371 € - 938 € (Selon les tarifs moyens d'une chambre standard)
Feb 28, 2016 - Réserver Altapura, Val Thorens sur TripAdvisor : consultez les 279 avis de voyageurs, 202 photos, et les meilleures offres pour Altapura, classé ...

Altapura (Val Thorens) : prix, photos et avis - France Hotel ...
www.france-hotel-guide.com/.../25648-altapura-val-th...
Situated at Val Thorens, the station de ski la plus élevée des Alpes, Val Thorens offers easy access to 600 km of ski slopes. This place is one ...

Accueil - Altapura : hôtel 5 étoiles Alpes Val Thorens 3 Vallées
www.altapura.fr/ Translate this page
Hôtel 5 étoiles à Val Thorens, plus haute station d'Europe au cœur du plus grand domaine skiable au monde, Altapura vous invite à laisser libre cours à vos ...

Altapura - Saint-Martin-de-Belleville | Reservation ... - Hotels
A major pitfall:
trust in online ratings decreases

- Representative surveys of French people
  - [Testntrust, 2013]
    - 89% trust online ratings in 2010
    - 76% trust online ratings in 2013
  - [Nielsen Institute, 2013]
    - 71% trust online ratings in 2007
    - 51% trust online ratings in 2013
Issues of online reputation metrics

- eBay
  - first to propose an online reputation solution in 1995
  - easier because
    - centralized
    - focused on one context only: online auctions
    - with real money transactions traces
- Issues
  - same points for successfully selling a Ferrari or a USB key
  - change in 2008: sellers cannot rate buyers in order to increase negative ratings of sellers
- aggressive marketing (Naymz/Visible.me spam, Reputation.com over alarming emails)
- reselling of private data without user consent (Rapleaf 1.0/Trustfuse)
- difficult and incomplete collection, verification and management of ratings
- TripAdvisor
  - Guilty of false ratings or successfully attacked
    - UK, 2009: sued by 2000 hotels association, change of slogan “reviews you can trust” to “reviews from our community”
    - France, 2011 : non-partner hotels listed as fully booked even if still available in real
    - Italy, 2014 and 2015:
      - fee of 500k Euros by the Italian anti-trust body due to unclear explanation regarding the validity of their ratings
      - ghost restaurant ranked as best restaurant of a touristic city
    - Tunisia, 2016: traveler's choice award given to the hotel in Tunisia where an Islamist terrorist attack left 30 British holidaymakers dead last summer
e-Reputation ratings main aspects

- Ratings verified or not
- Closed or open algorithms in order to evaluate their attack-resistance by the research community
  - security by obscurity is believed to be less secure by the research community
- Open, restricted or no API to access/manage them
- Their visualization or digital representation
  - Quantitative only
    • Scale of stars between 1 to 5...
  - Qualitative as well
    • Need of automated language sentiment analysis
How to visualize trust effectively?

• Trust visualization has a real business impact:
  +8% price premium [Johnston, 1996]
TrustPlus

- 2006 to 2012, decentralized, closed algorithm, not verified ratings, interesting trust visualization
• Score between 0 and 100
• Started in 2008
  – focusing on e-reputation influence
  – bought for around 100 millions $ in 2014
  – closed algorithm
  – based on detected evidence such as number of followers/fans and their own score engagement of posts
  – known to be easily attacked due to the easy set up of fake accounts
Fake Accounts, Clicks, Ratings and Reviews

Use our free tool as a guide to see if your followers are real, suspicious, or inactive. Simply type in any account name into the box below.

@checkmyaccount

0% suspicious or empty
0% inactive
0% good
Agenda

• Introduction
• Today’s Weak Online Reputation Metrics
• Computational Trust Engines
• Towards Standardized Attack-Resistant Trust Metrics
• Conclusion
• Q&A
Computational Trust

• One of its main goal is to achieve attack-resistant trust metrics
• A trust metric consists of the different computations and communications which are carried out by the trustor (and his/her network) to compute a trust value in the trustee
• A trust value is the digital representation of the trustworthiness or level of trust in the entity under consideration and is a non-enforceable estimate of the entity’s future behavior in a given context based on past evidence, mainly:
  – direct observations,
  – recommendations from an identified recommender,
  – reputation as an aggregated value from not clearly identified recommender(s).
• 3 main types of trust are considered in social research:
  – interpersonal trust,
  – dispositional trust,
  – system trust.
• Interpersonal trust is crucial when system trust cannot be enforced, for example, in the ubiquitous computing world of the Internet of Things (IoT).

[Seigneur, 2005]
McKnight & Cheverny
Trust Social Model
Trust Engine and Trust Metrics Attacks

- The trust metrics are attacked by means of:
  - Identity usurpation attacks
  - Identity multiplicity attacks
    - Douceur’s Sybil Attack is the most well-known
  - Coalitions of motivated users compared to other lazy users who do not rate
Research Representations of Trust Values

<table>
<thead>
<tr>
<th>Value range</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>Blind Trust</td>
</tr>
<tr>
<td>&gt; 0.9</td>
<td>Very high trust</td>
</tr>
<tr>
<td>0.75 to 0.9</td>
<td>High trust</td>
</tr>
<tr>
<td>0.5 to 0.75</td>
<td>High medium trust</td>
</tr>
<tr>
<td>0.25 to 0.5</td>
<td>Low medium trust</td>
</tr>
<tr>
<td>0 to 0.25</td>
<td>Low trust</td>
</tr>
<tr>
<td>-0.25 to 0</td>
<td>Low distrust</td>
</tr>
<tr>
<td>-0.5 to -0.25</td>
<td>Low medium distrust</td>
</tr>
<tr>
<td>-0.75 to -0.5</td>
<td>High medium distrust</td>
</tr>
<tr>
<td>-0.9 to -0.75</td>
<td>High distrust</td>
</tr>
<tr>
<td>&lt; -0.9</td>
<td>Very high distrust</td>
</tr>
<tr>
<td>-1</td>
<td>Complete distrust</td>
</tr>
</tbody>
</table>

[Marsh, 2016]

[SECURE, 2005]

[Wang and Vassileva, 2003]
Agenda

- Introduction
- Today’s Weak Online Reputation Metrics
- Computational Trust Engines
- Towards Standardized Attack-Resistant Trust Metrics
- Conclusion
- Q&A
Random Attack

4 randomly attacked
9 directly compromised
20 not compromised
Network Topology Engineered Attack

4 most connected attacked
20 compromised
9 not compromised
Trust Transfer: Sybil-attack Resistant Trust Metric

12 faked events may have been introduced in the network

[Seigneur, 2005]
The search for recommenders may be extended to contacts of recommenders. The total amount of trust transferred may be shared between several recommenders.

[Seigneur, 2005]
Conclusion

- Care must be taken when standardizing trust in order to not deceive the users and keep their trust in the trust standard.

- Attack-resistant trust metrics should be open and easy to be reviewed by the research community.

- Ideally, the most attack-resistant trust metrics should be standardized.
Q&A

• Thanks for your attention!
• Join the the 290+ Trustcomp community members
  – http://www.trustcomp.org/group-mailing-list
  – ACM SAC trust/reputation TRECK track CFP
  • Deadline: 15th September 2016

Jean-Marc.Seigneur@reputaction.com