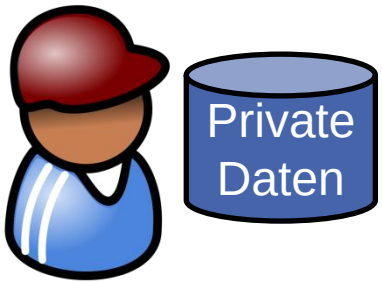
Karlsruher Institut für Technologie

Datenschutz bei notwendigen Veröffentlichungen privater Daten - Beispiele aus dem Gesundheits- und Energiebereich

Stephan Kessler

Institut für Programmstrukturen und Datenorganisation, Fakultät für Informatik, Karlsruher Institut für Technologie (KIT)



Vorstellung

- Mein Name
  - Stephan Kessler
  - stephan.kessler@kit.edu
- Doktorand am *Karlsruher Institut für Technologie* seit 2011
  - Stipendiat des Graduiertenkollegs **"Information Management and Market Engineering“**
  - Interdisziplinär: Wirtschaft, Informatik, Recht
- Forschung:
  - Datenschutz im Stromnetz der Zukunft
  - Datenschutzkonforme Auktionen auf einem lokalen Energiemarktplatz
    - Entwicklung von Simulationen mit Studenten

Datenschutzskandale – Ein Beispiel

Sie

Müll des Krankenhauses

Ihre Patientenakte

30.03.2012: **Patientenakten im Sperrmüll** taz.

Gemeinsamkeiten

- Veröffentlichung der Daten durch
  - Bedienfehler bzw. Unachtsamkeit
    - Facebook Partys
  - (mutwilliger) Verstoß gegen geltendes Recht
    - Patientenakten auf dem Sperrmüll

Bundesdatenschutzgesetz

- Zugang zu privaten Daten nach Bundesdatenschutzgesetz (BDSG):
  - Zugriff auf private Daten nur nach vorheriger Erlaubnis
  - Basiert auf Artikel 2 Grundgesetz:
    - „freie Entfaltung der Persönlichkeit“
- Risiko:
  - Ist Ihr Verhalten kontrollierbar, ändern Sie Ihr Verhalten

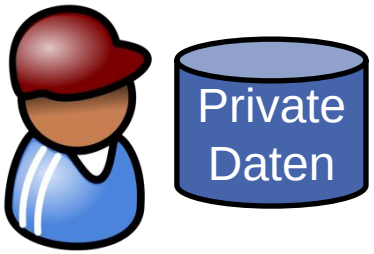
§

Veröffentlichung der Daten notwendig

Benötige eine große Menge Patientendaten um medizinische Forschungen durchzuführen!

Herausforderungen der Zukunft

- ***Individuen***:
  - Große Menge an privater, schützenswerter Daten.
  - Elektronisch gespeichert



- ***Gesellschaft***:
  - Daten benötigt um wichtige Ziele zu erreichen

- ***Ziele:***
  - Bereitstellung (‚Veröffentlichung') der Daten
  - Kein Eingriff in private Lebensbereiche (***Prävention***)

- *Wie können Daten bereitgestellt werden ohne Eingriff in die Privatheit einzelner?*

Vorgestellte Szenarien

	Gesundheitsdaten	Energiedaten
Betrifft	(fast) die gesamte Bevölkerung	(fast) die gesamte Bevölkerung
Informationen	Sensibel, persönlich	Sensibel, persönlich
Anwendungen	Medizinische Forschung	Energiewende, Emissionsreduktion

Agenda

- Motivation und Einleitung
- **Gesundheitsdaten**
  - Chancen bei Veröffentlichung
  - Risiken
  - Schutzmaßnahmen
- Energiedaten intelligenter Stromzähler (Smart Meter)
  - Intelligente Stromzähler und das Stromnetz der Zukunft
  - Eingriffe in die Privatheit
  - Ein lokaler Energiemarkt
    - Chancen für Energiewende & Risiken für Privatheit
    - Einfluss von Schutzmaßnahmen
- Fazit: Datenschutz & Prävention

Gesundheitsdaten – Chancen

- Illegale Giftmüll Deponien rund um Neapel
- Anstieg der Krebsraten um bis zu 300 %

- Verfügbarkeit von Daten nützlich:
  - Gesundheitsschutz

- ***Hier:***
  - statistische Daten ausreichend

- ***Generell:***
  - Datengranularität beliebig

Quelle: „Mafia entsorgt Millionen Tonnen Giftmüll in Italien“, Die Welt, 07.11.2013

Gesundheitsdaten - Risiken

Name	m/w	Geburt	PLZ	Gewicht	Krankheit
Paul Schmidt	m	30.07.1975	76131	50 kg	Bronchitis
Martin Mayer	m	13.10.1978	76351	150 kg	Angina
Nils Mustermann	m	17.01.1970	76131	48 kg	Erkältung
Annika Merkel	w	08.09.1987	68159	70 kg	Beinfraktur
Vanessa Müller	w	13.10.1980	68159	150 kg	Brustkrebs
Daniel Hoeneß	m	02.05.1964	10115	80 kg	Husten

- Gesundheitsdaten in tabellarischer Form
  - Zusatzinformationen notwendig
- Veröffentlichung ohne *Namen* und nur *Geburtsjahr*

Gesundheitsdaten - Risiken

m/w	Geburt	PLZ	Gewicht	Krankheit
m	1975	76131	50 kg	Bronchitis
m	1978	76351	150 kg	Angina
m	1970	76131	48 kg	Erkältung
w	1987	68159	70 kg	Beinfraktur
w	1980	68159	150 kg	Brustkrebs
m	1964	10115	80 kg	Husten

Re-Identifikation

- Veröffentlichung ohne *Namen* und nur *Geburtsjahr*

„Vanessa Müller“ ist im Datenbestand und ist stark übergewichtig! Welche Krankheit hat sie?

Angreifer

Dalenius, T. (1986). Finding a needle in a haystack-or identifying anonymous census record. Journal of official statistics, 2(3), 329-336.

Gesundheitsdaten – Risiken & Schutz

- Angreifer hat Hintergrundwissen
  - Interesse an Personen im näheren Umfeld
- Problem zuerst *1986* erkannt, *2002* gezeigt an Gesundheitsdaten eines Amerikanischen Krankenhauses in Kombination mit der Wählerdatei
- Einfachste Schutzmaßnahme: Veränderung der Daten
  - *Identifikatoren:* **Name/Anschrift**
    - Entfernen
  - *Quasi-Identifikatoren:* **Geschlecht, Gewicht, PLZ, etc.**
    - Zu Gruppen von *k*-Elementen zusammenfügen
  - *Sensitives Attribut:* **Krankheit**
    - erhalten

Angreifer

Gesundheitsdaten - Schutzmaßnahmen

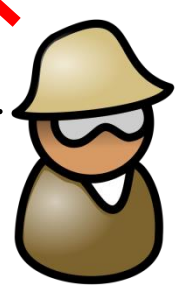
Quasi-Identifikatoren

m/w	Geburt	PLZ	Gewicht	Krankheit
m	197*	76***	~100 kg	Bronchitis
m	197*	76***	~100 kg	Angina
*	19**	*****	~59 kg	Erkältung
*	19**	*****	~59 kg	Beinfraktur
*	19**	*****	~115 kg	Brustkrebs
*	19**	*****	~115 kg	Husten

Re-Identifikation

- Veröffentlichung in Gruppen aus zwei Zeilen

„Vanessa Müller“ ist im Datenbestand und ist stark übergewichtig! Welche Krankheit hat sie?



Angreifer

- *k-Anonymität (k=2)*

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(05), 557-570.

Gesundheitsdaten - Fazit

m/w	Geburt	PLZ	Gewicht	Krankheit
*	19**	*****	~115 kg	Brustkrebs
*	19**	*****	~115 kg	Husten

- Ergebnis stark verfälscht
  - Sind die Daten noch nützlich? (z.B. zur Forschung?)
  - Sind vielleicht andere Gruppierungen „besser“
- Privatheit nicht garantiert:
  - Angreifer weiß, es handelt sich um einen Mann
  - *Brustkrebs* ist also auszuschließen

Angreifer

Gesundheitsdaten - Fazit

- Wichtige Erkenntnisse
  - Anonymisieren von Daten ***nicht trivial***
  - Mögliche Angriffe vielfältig
  - Änderungen an den Daten nötig
  - Resultierende Datenqualität schwer zu bewerten

- Nächster Schritt:
  - Betrachtung eines Szenarios im Stromnetz der Zukunft
  - Bewertung der Datenqualität

Angreifer

Vorgestellte Szenarien

	Gesundheitsdaten	Energiedaten
Betrifft	(fast) die gesamte Bevölkerung	(fast) die gesamte Bevölkerung
Informationen	Sensibel, persönlich	Sensibel, persönlich
Anwendungen	Medizinische Forschung	Energiewende, Emissionsreduktion
Datenqualität	**Schwer zu bewerten**	**Quantifizierbar (€, Tonnen)**

Agenda

- Motivation und Einleitung
- Gesundheitsdaten
  - Chancen bei Veröffentlichung
  - Risiken
  - Schutzmaßnahmen
- **Energiedaten intelligenter Stromzähler (Smart Meter)**
  - Intelligente Stromzähler und das Stromnetz der Zukunft
  - Eingriffe in die Privatheit
  - Ein lokaler Energiemarkt
    - Chancen für Energiewende & Risiken für Privatheit
    - Einfluss von Schutzmaßnahmen
- Fazit

Stromnetz der Zukunft

- Aktuelles Stromnetz (‚althergebracht‘)
- Schlechte Einbindung regenerativer Energien
- Klimaziele (CO_2 Reduktion) schwer zu erreichen

- Deshalb: Stromnetz der Zukunft – Smart Grid
  - Einsatz von Kommunikationstechnik um Informationsaustausch zu beschleunigen

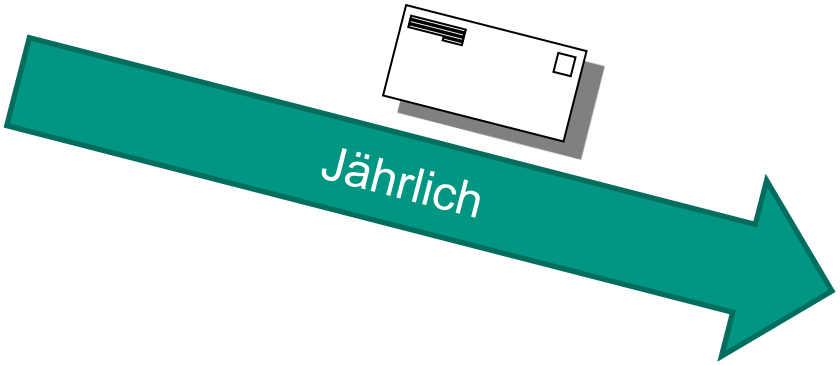
Ferraris Zähler

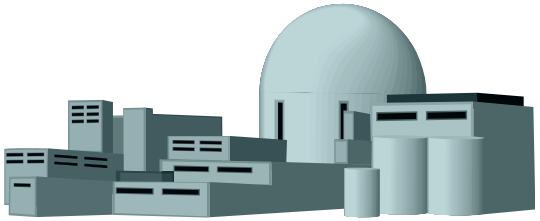
Smart Meter

Intelligente Stromzähler – Funktionsweise

Ferraris Zähler

Smart Meter





Intelligente Stromzähler – Nutzen

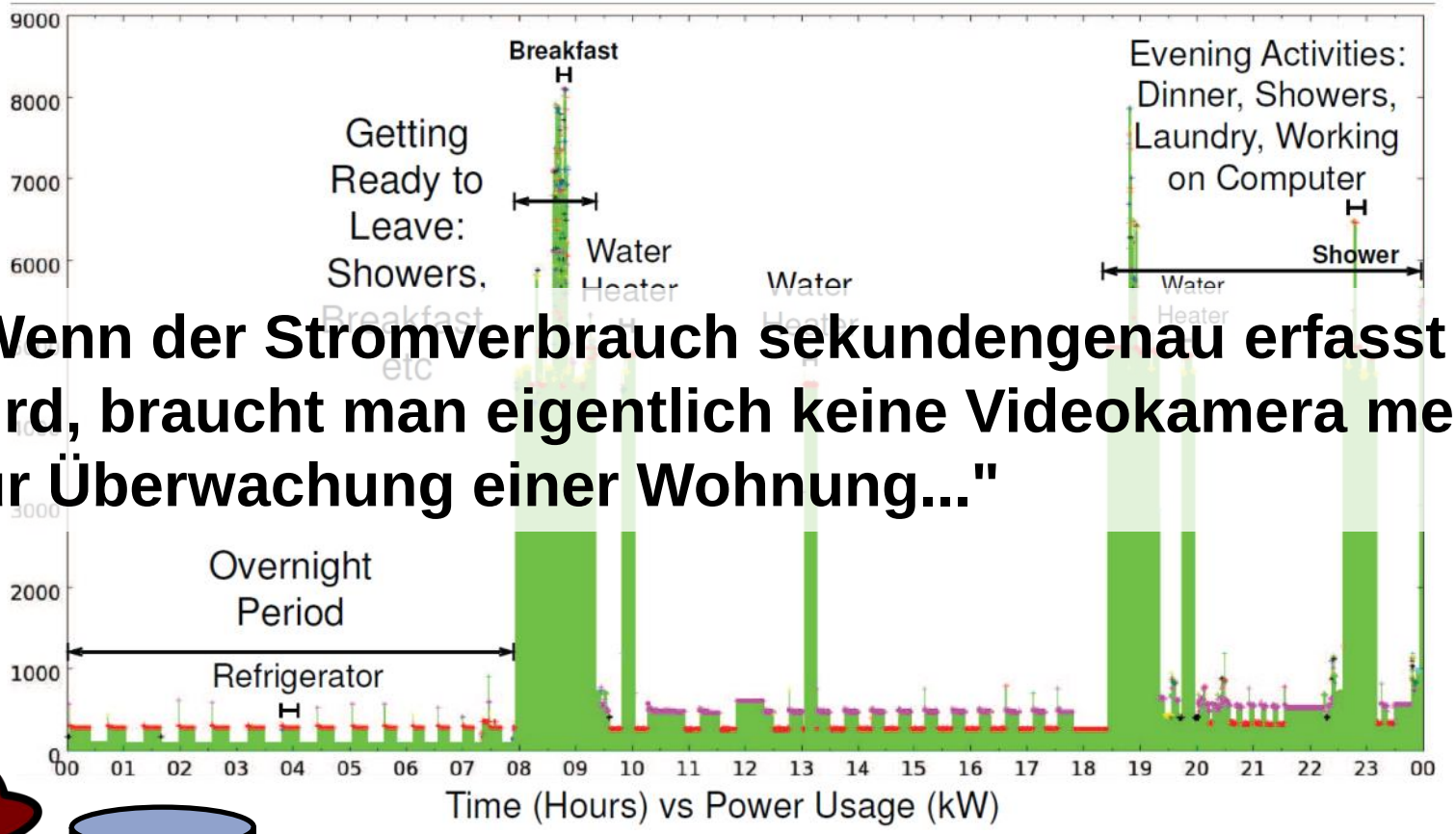
- Genauere Informationen über Verbrauchsmuster
- Verbesserung von
  - Produktionsplanung
  - Netzplanung
- Lokaler Energiemarkt

- Zusätzliche Möglichkeiten
  - Ferngesteuertes ein/ausschalten von Verbrauchern
  - Lademanagement von Elektroautos
  - …

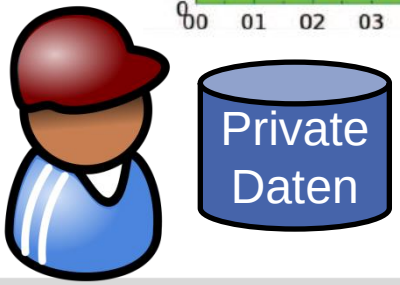
- Nachteil:
  - Daten enthalten viele private Informationen

Smart Meter

Intelligente Stromzähler – Information in Daten



"Wenn der Stromverbrauch sekundengenau erfasst wird, braucht man eigentlich keine Videokamera mehr zur Überwachung einer Wohnung..."



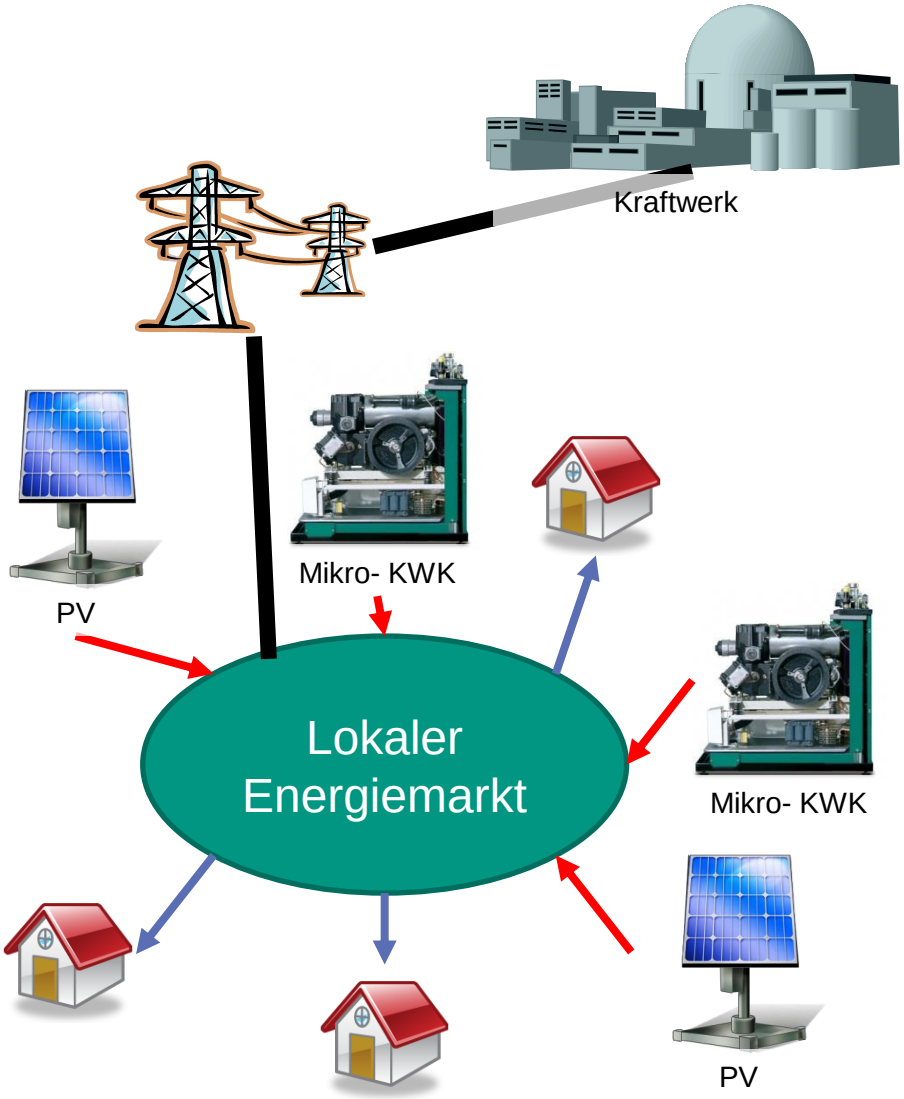
A. Molina-Markham et al., "Private Memoirs of a Smart Meter," 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency, 2010

Agenda

- Motivation und Einleitung
- Gesundheitsdaten
  - Chancen bei Veröffentlichung
  - Risiken
  - Schutzmaßnahmen
- Daten intelligenter Stromzähler (Smart Meter)
  - Intelligente Stromzähler und das Stromnetz der Zukunft
  - Eingriffe in die Privatheit
  - **Ein lokaler Energiemarkt**
    - Chancen für Energiewende & Risiken für Privatheit
    - Einfluss von Schutzmaßnahmen
- Fazit: Datenschutz & Prävention

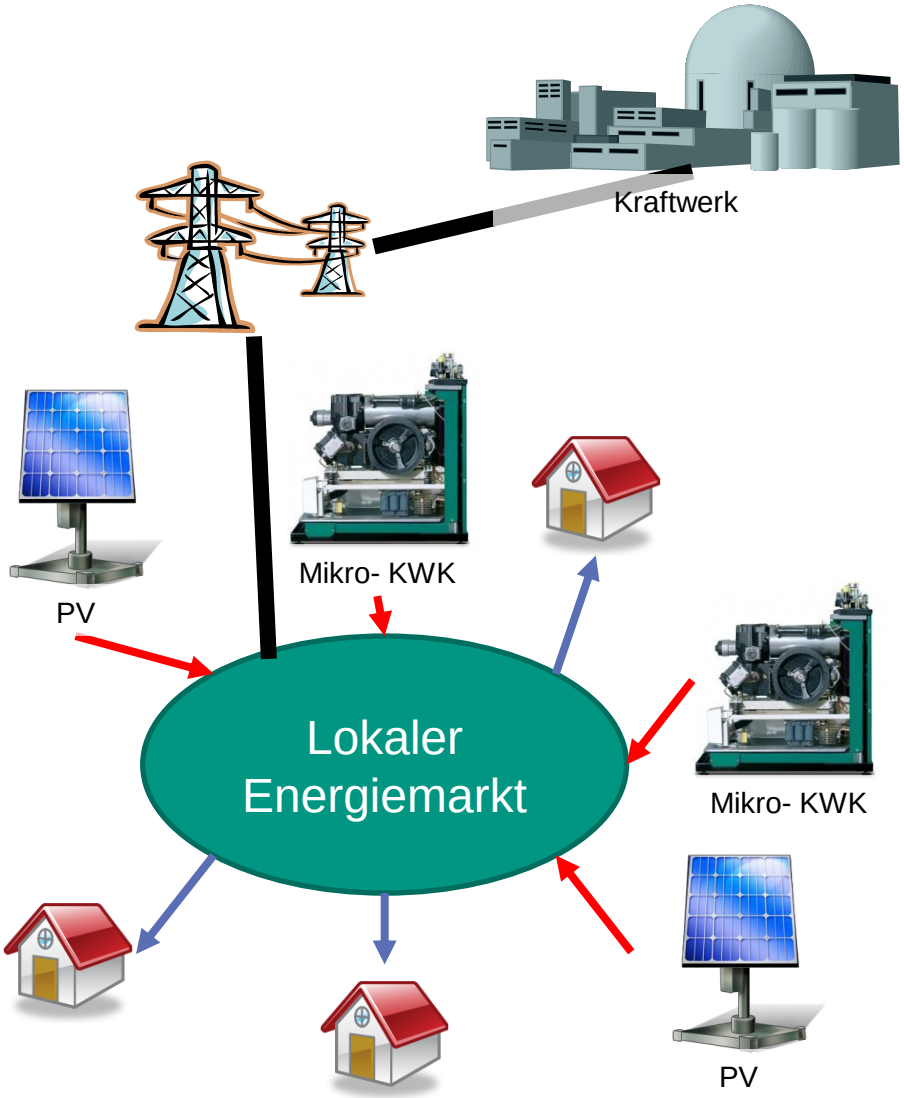
Lokaler Energiemarkt

- Populäre Regenerative Quellen
  - Photovoltaik (PV)
  - Kraft-Wärme-Kopplung (KWK)
- Probleme:
  - Verlust durch Transport über lange Leitungswege
  - Koordination von Erzeugung und Verbrauch



Lokaler Energiemarkt

- Lösung:
  - Energiemarktplatz für einen Ort
  - Handel im 15-Minuten Takt
- *Chancen:*
  - *Reduktion von CO_2 Emissionen*
  - *Effiziente PV/KWK Nutzung*

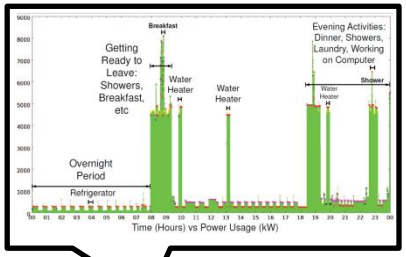


Buchmann, E., Kessler, S., Jochem, P., & Böhm, K. (2013). The Costs of Privacy in Local Energy Markets. 2013 IEEE 15th Conference on Business Informatics, 198–207.

Lokaler Energiemarkt

- Gebote der Haushalte entsprechen der privaten Daten
  - Tatsächlicher Verbrauch möglichst von regenerativen Energien gedeckt

- Möglichkeiten zum Schutz der Privatheit:
  - **Daten modifizieren und „falsche“ Gebote abgeben**



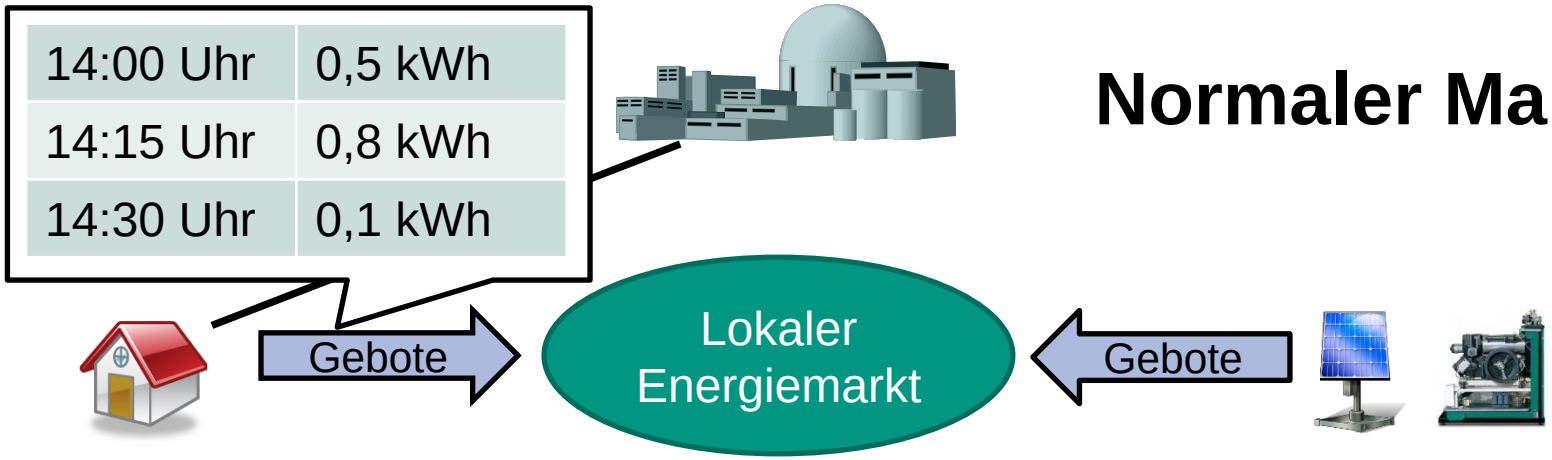
Gebote

Lokaler Energiemarkt

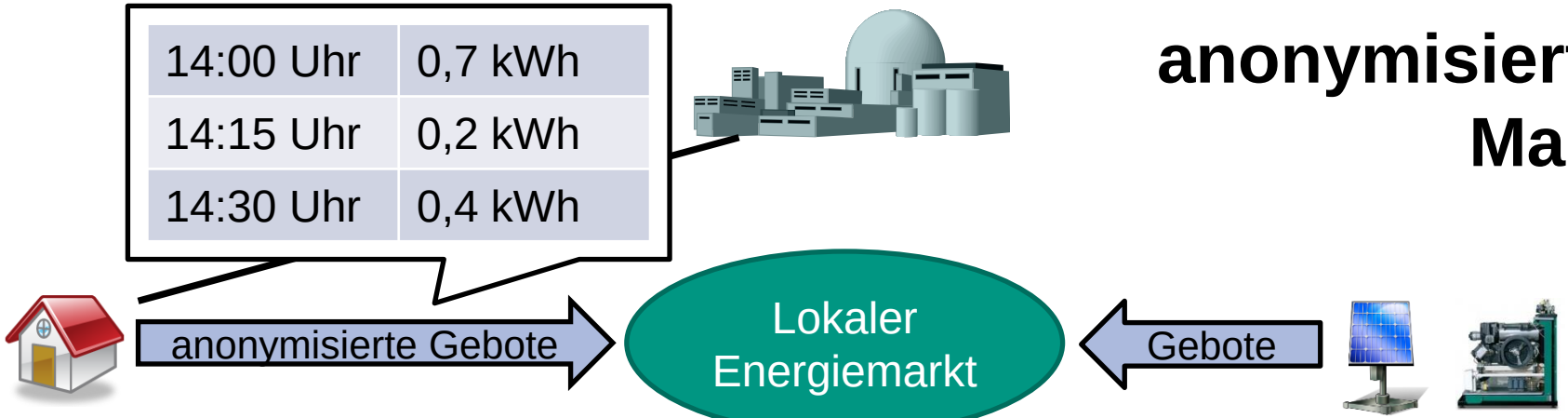
Gebote

KAUF	VERKAUF
14:00 – 14:15 Uhr 0,2 kwh 20 Cent/kwh	14:00 – 14:15 Uhr 1,2 kwh 15 Cent/kwh

Lokaler Energiemarkt



Normaler Markt



anonymisierter Markt

Lokaler Energiemarkt

- Abgabe anonymisierter Gebote
  - Zusatzkosten für Haushalte
  - Schlechtere CO_2 Effizienz

- Einfluss der Anonymisierung auf Anwendung messbar

Lokaler Energiemarkt

- Simulation einer Ortschaft mit 5000 Haushalten
  - Ca. 7500 Bewohner
  - Unterschiedliche Anzahl von PV und KWK Anlagen
  - Unterschiedliche starke Anonymisierungsverfahren

	Normaler Markt	Anonymisierter Markt
Kosten für Elektrizität	Ø 11,00 € / Woche	Ø 13,10 € / Woche
CO_2 Einsparungen	~ 110 Tonnen	~ 105 Tonnen

Lokaler Energiemarkt - Erkenntnisse

- Daten Intelligenter Stromzähler enthalten persönliche Informationen
- Veröffentlichung notwendig (z.B. Energiemarkt) um:
  - CO_2 Emissionen einzusparen
  - Funktionalität des Stromnetzes sicherzustellen

- Einfluss von Schutzmaßnahmen
  - Messbar
  - Nicht signifikant

Agenda

- Motivation und Einleitung
- Gesundheitsdaten
  - Chancen bei Veröffentlichung
  - Risiken
  - Schutzmaßnahmen
- Energiedaten intelligenter Stromzähler (Smart Meter)
  - Intelligente Stromzähler und das Stromnetz der Zukunft
  - Eingriffe in die Privatheit
  - Ein lokaler Energiemarkt
    - Chancen für Energiewende & Risiken für Privatheit
    - Einfluss von Schutzmaßnahmen
- **Fazit: Datenschutz & Prävention**

Fazit: Datenschutz & Prävention

- Datenschutzskandale im Bewusstsein der breiten Bevölkerung
- Herausforderung der Zukunft:
  - Veröffentlichung von Daten aufgrund gesellschaftlichen Nutzens
  - Anonymisierung der Daten nicht trivial
- Forschung arbeitet an *Kompromiss aus Privatheit und Nutzen*
  - Ergebnisse vielversprechend
- ***Prävention vor Eingriffen in die Privatheit***
  - ***Betrifft ein Großteil der Bevölkerung***