Vertrauen in einer digitalisierten Welt

Helmut Schweinzer
4. Mai 2024
Linuxtag der Erlanger Linux User Group

Überblick

Vertrauen allgemein:

- Wahrnehmung
- Was ist denn Vertrauen?
- Woher bekommt man Vertrauen?
- Wozu braucht man Vertrauen?
- Wie verliert man Vertrauen?

In der IT:

- Aspekte von Vertrauen in der IT
- Vertrauens-Techniken
- Malware
- Künstliche Intelligenz
- DeepFakes
- Linux

Allgemeines Vertrauen

Komplexe Welt

Wahrnehmung

Was war das Wichtigste?

Wie funktioniert Wahrnehmung?

- Die menschliche “Software” vereinfacht
  → das Maus-Experiment
- individuelle, erlernte “Filter”
  verringern Komplexität
- Bekanntes, Beherrschbares und Unwichtiges wird ausgeblendet

Definition von Vertrauen

Vertrauen ist die Annahme, dass sich eine bestimmte Angelegenheit in die (für mich) richtige Richtung bewegt.

- Es geht dabei um eine für mich genügend hohe Wahrscheinlichkeit
- Vertrauen verringert Komplexität
- Was dann noch fehlt ist ungewiss und erzeugt Stress.

Komplexitätsverringerung

Wer alles weiß braucht nicht zu vertrauen

Einfachheit und Vertrauen

- Vertrauen schwindet, wenn die Welt durch diese Mechanismen nicht soviel einfacher wird, wie erhofft
- Die Welt ändert sich ständig
- “Früher war alles besser”
  - Logisch! Mit dem Wissen von heute
  - nützt nichts – die Zeit läuft vorwärts

Grundlagen: Der Andere

- was er mir sagt
- was er anderen sagt
- was er schreibt
- was er tut
- was er tun will
- was er erreichen will
- wie konsistent sein Verhalten ist
- wie aufrichtig er mir gegenüber ist
- wie er mein Vertrauen bisher bestätigt hat
- wie sehr er mir vertraut
- …

Grundlagen: Ich

- was ich gehört habe
- was ich gelesen habe
- was ich gesehen habe
- was ich meine, dass der andere will/soll
- was ich sicher weiß
- was ich meine zu wissen
- was ich vermutet habe
- wie oft meine Vermutungen zugetroffen haben
- was ich erreichen will
- was ich tun muss
- …

Andere Grundlagen

- Grundvertrauen
- das Gute im Menschen
- Lebenserfahrung, also ein längerer Prozess
- Selbstvertrauen
- “Alle” machen es so (Quantität)
- Vertrauensvorschuss
  - Kompetenzerwartung
  - Integritätserwartung
  - Benevolenzerwartung
- … Naivität

Aspekte von Vertrauen

- Persönliches Vertrauen ist subjektiv
- Digitales Vertrauen ist objektiv … oder?

- Relevanz: Was passiert, wenn mein Vertrauen nicht gerechtfertigt war?

- “Vertrauen ist gut, Kontrolle besser”
- Diktaturen leben vom Misstrauen

Wo wird Vertrauen gebraucht

- Persönliche Beziehungen / andere Menschen
- Politik
- Technik
  - Robustheit / Stabilität
  - Computer (die machen alles richtig - müssen sie ja, weil deterministisch)
  - Künstliche Intelligenz
- Information
  - Richtigkeit von etwas
  - Nachrichten (Fakes)
  - Bilder, Videos
- Wissenschaft, Medizin, ...
- Abschreckung, Überlegenheit, Macht des Geldes ...

Arten des Vertrauens

Gerichtetes Vertrauen:

A → B

Bidirektionales Vertrauen:

A ↔ B

Indirektes Vertrauen:

A → B → C

Nahes Vertrauen (jetzt):

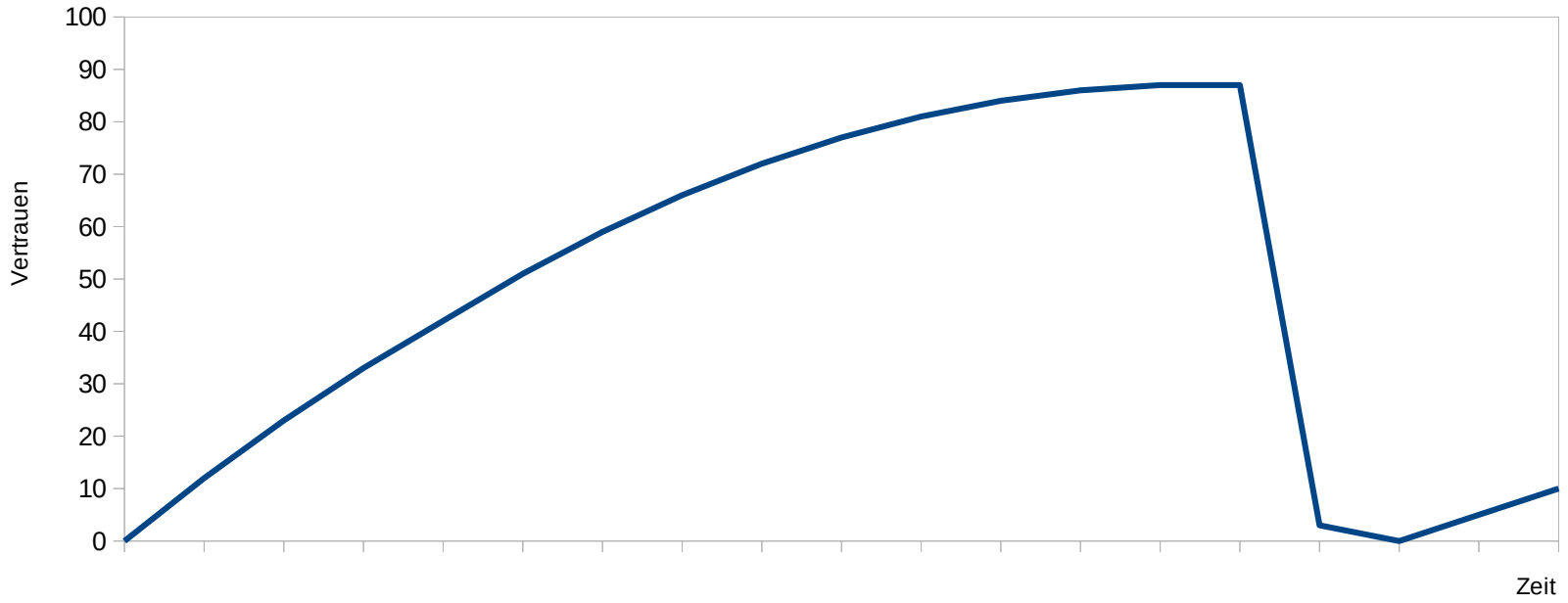
A → B

Fernes Vertrauen (später):

A - - - - - → B

Vertrauensverlust

- Vertrauensverlust
- Misstrauen



- Vertrauen zurückgewinnen dauert …

Vertrauen in der IT

Digitalisierte Welt

Vertrauen in der IT

- IT soll funktionieren
- Integrität (meines Computers, meiner Daten, meiner "Partner")
- Reproduzierbarkeit
- Fehlerfreiheit (von Hardware, Programmen, Systemen)
- Sicherheit
- Vertrauenswürdigkeit der “Daten”

Vertrauen bei digitaler Kommunikation

- mein Gegenüber ist echt (kein Bot)
- meine Daten kommen unverfälscht beim Partner an
- die Daten kommen unverfälscht von meinem Partner bei mir an
- niemand manipuliert mich

Vertrauen zu meinem Computer

- dass die Hardware funktioniert
- dass das System so bootet wie immer
- dass das OS vertrauenswürdig ist
- dass das OS sicher gepatcht wird
- dass die Software das tut, was sie soll / was ich will
- dass meine Daten mir gehören …
- dass meine Daten erhalten bleiben → Backup

Digitales Vertrauen

- Hashes
- Verschlüsselung
- Signaturen
- Zertifikate
  - CAs – “Authorities” – Institutionen deren Geschäft das Vertrauen ist
  - Web-of-Trust (PGP)
- Trusted-Boot, Anti-Evil-Maid (TPM)
- Blockchain
- …

Malware und Vertrauen

- Malware/Phishing versucht Vertrauen aufzubauen:
  - das ist gerade aktuell
  - andere Experten … meinen auch, dass …
  - ich kümmere mich um dich
    - ich gebe dir Tipps: du solltest …
    - ich warne dich vor …
    - Du könntest etwas Lukratives erreichen …
- Vertrauen aufzubauen dauert lange
  → Zeitdruck machen, da das bisherige Vertrauen reichen muss.

Künstliche Intelligenz

Um Vertrauen in eine Technologie haben zu können, muss man (zumindest ansatzweise) die Wirkungsweise verstehen und die Risiken kennen

AI / KI

- Mustererkennung
- Neuronale Netze
- LLM
- GPT
- Generative Adversarial Networks
- Clustering
- Moments
- ...

Probleme der KI/AI

Richtigkeit der Daten:

- Daten-Bias
- Fehlerhafte Daten
- Zu kleine Menge an Trainingsdaten
- Mengen verdecken Details
- Unvollständige Daten → Raten
- Intransparenz
- nur Wahrscheinlichkeit von Wörtern
- Nur gesammelte Daten – keine Kreativität
- …

Probleme der KI/AI

Anwendungen:

- Manipulation:
  - Fake News
  - Deep Fakes
  - Überwachung
  - Social Media Fake Accounts
- Betrug
- Wahlen
- Krieg

FakeNews erkennen

- Reißerisch und mit vielen unbewiesenen Behauptungen
- Gibt es Quellenangaben und sind die plausibel?
- Ist der Wortlaut identisch? – Wo ist die Quelle?
- Bei Fotos Quelle z.B. mit Google-Bildsuche suchen
- Bilder-Metadaten anschauen
- …

DeepFakes erkennen

- Technisch:
  - glatte Kanten / unterschiedliche Schärfe
  - falsche Lichtrichtung
  - unpassende Details (nicht nur Gesicht sondern auch Hände, Haare, Körperform, ...)
  - unnatürliche Bewegungen
  - Synchronizität Stimme - Bild
  - falsche Stimme
  - Passt das Wetter im Video zum aktuellen?
  - Stimmt die Umgebung (z.B StreetView)?

DeepFakes erkennen

- Inhalt:
  - Unplausibler Inhalt (kann das stimmen?)
  - Versucht das Video zu manipulieren, beunruhigen, Angst zu machen?
  - Gibt es andere Veröffentlichungen zu dem Thema (die sich nicht auf das Video beziehen oder es als Quelle verwenden)?
  - Gibt es Videos mit der entgegengesetzten Botschaft?
  - Wird diese Nachricht auch in von dir vertrauten Medien gebracht? (begündetes, detailliert erarbeitetes Vertrauen)
  - Augenzeugen/Experten … befragen
  - Selber hingehen

Software zur DeepFake-Erkennung

- Software
  - Sentinel
  - Intel: FakeCatcher
  - Microsoft: Video Authenticator
  - WeVerify
  - …
- “Moreover, it's important to remember that technology alone cannot solve the problem of deepfakes.”
- Regulierungen und Gesetze (Wasserzeichen, …) funktionieren nicht.

Vertrauen: OSS und Linux

- Transparenz (Open Source)
- Security by Design
- Stabilität (→ Server)
- Performance
- Kompatibilität (offene Standards)
- Viele Augen schauen drauf (→ schnelle Bugfixes)
- Anpassbar und erweiterbar (auch härtbar)
- LTS
- Community Support
  - nicht zuletzt durch die ERLUG ;-)
- Innovativ

Danke für die Aufmerksamkeit

Fragen ?

Helmut Schweinzer
Email: hel@ki-aikido.de
PGP: 0x19851166
Fingerprint: C6D3 B58F 7E3A C251 A5C5 2F89 E377 2547 1985 1166