(19) Deutsches Patent- und Markenamt

(10) **DE 10 2017 207 574 A1** 2018.11.08

(12) # Offenlegungsschrift

(21) Aktenzeichen: **10 2017 207 574.0**
(22) Anmeldetag: **05.05.2017**
(43) Offenlegungstag: **08.11.2018**

(51) Int Cl.: ***G06F 17/20*** (2006.01)

(71) Anmelder:
Siemens Aktiengesellschaft, 80333 München, DE

(72) Erfinder:
Langen, Manfred, 81827 München, DE

(56) Ermittelter Stand der Technik:

US 8 489 635 B1

AHMED, Ahmed A.; TRAORE, Issa. Biometric recognition based on free-text keystroke dynamics. IEEE transactions on cybernetics, 2014, 44. Jg., Nr. 4, S. 458-472. doi: 10.1109/TCYB.2013.2257745

SHOUNAK, Devbhuti, et al. A Method for Bypassing Keystroke Recognition Based Security System Using Social Engineering. 2014. IOSR-JCE, ISSN: 2278-8727, Volume 16, Issue 2, PP 87-93, Mar-Apr. 2014.

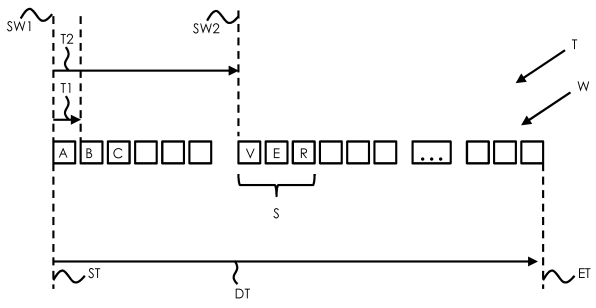
STEFAN, Deian; SHU, Xiaokui; YAO, Danfeng Daphne. Robustness of keystroke-dynamics based biometrics against synthetic forgeries. computers & security, 2012, 31. Jg., Nr. 1, S. 109-121. doi: 10.1016/j.cose.2011.10.001

Prüfungsantrag gemäß § 44 PatG ist gestellt.

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **Verfahren zur Erkennung eines maschinengenerierten Textes sowie Verfahren zu ihrer Vereitelung**

(57) Zusammenfassung: Bei dem Verfahren zur Erkennung eines maschinengenerierten Textes in einem Netzwerkforum wird der zeitliche Verlaufeiner Eingabe des Textes zur Erkennung herangezogen.
Bei dem Verfahren zur Vereitelung einer Erkennung eines maschinengenerierten Textes in einem Netzwerkforum wird der zeitliche Verlauf einer Eingabe des Textes zur Vereitelung der Erkennung herangezogen.



Beschreibung

[0001] Die Erfindung betrifft ein Verfahren zur Erkennung eines maschinengenerierten Textes in einem Netzwerkforum sowie ein Verfahren zur Vereitelung einer Erkennung eines maschinengenerierten Textes in einem Netzwerkforum.

[0002] In jüngerer Zeit kommen vermehrt Chatbots auf, welche computergeneriert Beiträge in Netzwerkforen einstellen. Es ist bekannt, maschinengenerierte Texte von Chatbots aufgrund der Syntax, der Semantik und der Rechtschreibung zu erkennen. Solche Lösungen sind jedoch sehr aufwändig.

[0003] Es ist daher Aufgabe der Erfindung, ein verbessertes Verfahren zur Erkennung eines maschinengenerierten Textes in einem Netzwerkforum anzugeben, welches insbesondere einfacher als bislang bekannt ausgeführt werden kann. Es ist ferner Aufgabe der Erfindung, ein Verfahren zur Vereitelung einer solchen Erkennung eines maschinengenerierten Textes in einem Netzwerkforum anzugeben. Das heißt, es soll zugleich auch ein Verfahren angegeben werden, mit welchem das erfindungsgemäße Verfahren zur Erkennung eines maschinengenerierten Textes in einem Netzwerkforum unterlaufen werden kann.

[0004] Diese Aufgabe der Erfindung wird mit einem Verfahren zur Erkennung eines maschinengenerierten Textes in einem Netzwerkforum mit den in Anspruch 1 angegebenen Merkmalen sowie mit einem Verfahren zur Vereitelung einer Erkennung eines maschinengenerierten Textes in einem Netzwerkforum mit den in Anspruch 2 angegebenen Merkmalen gelöst. Bevorzugte Weiterbildungen der Erfindung sind in den Unteransprüchen, der nachfolgenden Beschreibung und der Zeichnung angegeben.

[0005] Bei dem erfindungsgemäßen Verfahren zur Erkennung eines maschinengenerierten Textes in einem Netzwerkforum wird der zeitliche Verlauf einer Eingabe des Textes zur Erkennung erfasst und herangezogen. Der zeitliche Verlauf der Eingabe des Textes bildet neben dem Inhalt des Textes selbst einen eigenen Parameterraum, in welchem sich die zeitlichen Verläufe der Eingaben von maschinengenerierten Texten typischerweise deutlich von jenen zeitlichen Verläufen von Eingaben von Texten durch menschliche Nutzer unterscheiden. Erfindungsgemäß wird daher dieser weitere Parameterraum zur Erkennung maschinengenerierter Texte herangezogen.

[0006] Bei dem erfindungsgemäßen Verfahren zur Vereitelung einer Erkennung eines maschinengenerierten Textes in einem Netzwerkforum wird der zeitliche Verlauf einer Eingabe des Textes zur Vereitelung der Erkennung herangezogen, d.h. genutzt. Das bedeutet, es wird der zeitliche Verlauf der - maschinengenerierten - Eingabe des Textes dem zeitlichen Verlauf einer Eingabe des Textes durch menschliche Nutzer angeglichen, so dass der zeitliche Verlauf der Eingabe des Textes gemäß diesem Aspekt der Erfindung keine zur Erkennung eines maschinengenerierten Textes förderlichen Hinweise bietet.

[0007] Vorzugsweise ist bei den erfindungsgemäßen Verfahren das Netzwerkforum ein Internetforum.

[0008] Bevorzugt umfasst bei den erfindungsgemäßen Verfahren der zeitliche Verlauf der Eingabe des Textes zumindest die Gesamtzeitdauer der Eingabe des Textes. So ist typischerweise die Gesamtdauer für die Einstellung eines Beitrags in einem Netzwerkforum im einfachsten Falle eines computergenerierten Textes äußerst kurz, da der gesamte Text als vollständiger Block in das Netzwerkforum eingetragen, etwa schlicht einkopiert, wird. Folglich würde eine derart geringe Zeit für die Abfassung des gesamten Textes erfasst werden, dass diese geringe Zeit mit nahezu vollständiger Sicherheit auf einen maschinengenerierten Text hinweist. Erfindungsgemäß vereitelt wird die auf der Gesamtzeitdauer der Eingabe des Textes basierende Erkennung eines maschinengenerierten Textes mittels einer Anpassung der Gesamtzeitdauer bei der Eingabe an eine solche Gesamtzeitdauer, welche einer typischen Gesamtzeitdauer der Eingabe des Textes durch menschliche Nutzer entspricht.

[0009] In einer vorteilhaften Weiterbildung der erfindungsgemäßen Verfahren umfasst der zeitliche Verlauf die Zeitdauer der Eingabe zumindest eines Wortes und/oder zumindest eines Zeichens und/oder Buchstabens des Textes. Solche Zeitdauern lassen sich zum einen rechnerisch durch Division der Gesamtzeitdauer des Textes durch die Länge des Textes in Wörtern und/oder Zeichen und/oder Buchstaben ermitteln. Alternativ oder zusätzlich lässt sich die Zeitdauer der Eingabe eines Wortes und/oder eines Zeichens und/oder Buchstabens mittels nutzerseitig installierter Programme, insbesondere mittels eines Scripts, erfassen, wie es beispielsweise von der eingabesynchronen Autovervollständigung von Eingaben in Eingabemasken von Suchmaschinen bekannt ist. Auf diese Weise lassen sich solche Zeitdauern für die erfindungsgemäßen Verfahren zur Erkennung nutzen oder aber zur Vereitelung dieser Erkennung manipulieren.

[0010] In einer vorteilhaften Weiterbildung der Verfahren gemäß der Erfindung umfasst der zeitliche Verlauf die Zeitdauer der Eingabe zumindest einer Silbe oder zumindest einer Zeichenkombination, insbesondere einer Buchstabenkombination. Insbesondere können Zeichen- und/oder Buchstabenkombinationen kurze, vorzugsweise häufig verwendete, Wörter oder Silben bilden. Gerade bestimm-

te Silben oder Buchstabenkombinationen weisen bei der Eingabe durch Menschen typische Muster auf. Beispielsweise ist ein solches Muster eine aus der Tippgewohnheit heraus resultierende, vergleichsweise rasche, Eingabe kurzer Wörter wie etwa „der“, „die“, „das“ und von Silben wie (im Deutschen) „ver-“, „gen-“, „-den“. Eine raschere Eingabe solcher Zeichen- und/oder Buchstabenkombinationen und Silben im Vergleich zu weniger geläufigen Silben oder Zeichen- und/oder Buchstabenkombinationen bildet folglich ein Indiz für einen Text, welcher durch einen menschlichen Nutzer generiert ist, so dass eine Erkennung gemäß der Erfindung weniger stark auf einen maschinengenerierten Text hinweisen wird. Zugleich kann eine Vereitelung einer solchen Erkennung mittels Anpassung der Zeitdauer der Eingabe solcher Silben oder Zeichen- und/oder Buchstabenkombinationen an menschliche Nutzer erfolgen.

[0011] Idealerweise umfasst bei den erfindungsgemäßen Verfahren der zeitliche Verlauf ein Maß für die Streuung der Zeitdauer, wie vorhergehend erläutert. So kann beispielsweise bei der Eingabe einzelner Wörter oder Zeichen- und/oder Buchstabenkombinationen eine Erkennung unzuverlässig sein, indem beispielsweise ein Text zwar durch einen menschlichen Nutzer generiert wird, aber Teile von Texten oder Passagen, insbesondere bei Zitaten, bei der Eingabe computerbasiert kopiert und eingefügt werden. In diesen Fällen würden bei einer isolierten Betrachtung von Zeitdauern für die Eingabe eines Wortes falschpositive Erkennungen maschinengenerierter Texte erfolgen. Indem die Streuung der Zeitdauer erfasst wird, können solche computerunterstützten Spezialfälle bei der Eingabe von Wörtern oder Zeichen- und/oder Buchstabenkombinationen berücksichtigt werden.

[0012] Insbesondere wird bei den erfindungsgemäßen Verfahren bei dem zeitlichen Verlauf ein Anteil kopierten Textes berücksichtigt. Auch in dieser Weiterbildung der Erfindung wird berücksichtigt, dass selbst bei einer Eingabe eines Textes durch einen menschlichen Nutzer einzelne Wörter oder Textpassagen als Block eingefügt werden können.

[0013] Besonders bevorzugt umfasst bei den erfindungsgemäßen Verfahren der zeitliche Verlauf den zeitlichen Verlauf der Eingabe von nebeneinanderliegenden Zeichen und/oder Buchstaben und/oder nicht mit derselben Hand zu betätigenden Zeichen und/oder Buchstaben auf einer Normtastatur, insbesondere einer QWERTZ- oder QWERTY-Tastatur oder einer Dvorak-Tastatur. Je nach Tastatur und Tippgewohnheiten von menschlichen Nutzern werden nebeneinanderliegende Zeichen und/oder Buchstaben und/oder nicht mit derselben Hand zu betätigende Zeichen und/oder Buchstaben auf einer Normtastatur in einer zeitlichen Abfolge betätigt, deren zeitlicher Verlauf charakteristische Auffälligkeiten aufweist. So werden häufig bei einem wenig geübten Schreiber nebeneinander auf der Tastatur liegende Zeichen und/oder Buchstaben relativ schnell aufeinanderfolgend betätigt, während häufig weit entfernt liegende Buchstaben weniger schnell getippt werden können. Umgekehrt werden etwa bei geübten Schreibern häufig mit verschiedenen Händen betätigte Zeichen und/oder Buchstaben besonders schnell aufeinanderfolgend geschrieben. Anhand dieser charakteristischen Auffälligkeiten können einerseits maschinengenerierte Texte erkannt werden. Andererseits können solche Auffälligkeiten bewusst bei einer maschinengenerierten Abfassung von Texten genutzt werden, so dass eine Erkennung eines maschinengenerierten Textes vereitelt ist.

[0014] Besonders bevorzugt wird bei den erfindungsgemäßen Verfahren der zeitliche Verlauf mit einem Referenzverlauf verglichen. Ein solcher Referenzverlauf kann beispielsweise aus dem Betrieb eines Netzwerkforums selbst erhalten werden, indem beim Betrieb dieses Netzwerkforums der zeitliche Verlauf der Eingaben von Texten in diesem Netzwerkforum erfasst und statistisch ausgewertet wird. Insbesondere können Mittelwerte und Standardabweichungen von für den zeitlichen Verlauf der Eingabe charakteristischen Größen erfasst und buchgehalten und für einen Vergleich herangezogen werden. Vorzugsweise wird bei den erfindungsgemäßen Verfahren der Referenzverlauf mittels eines Referenznetzwerkforums ermittelt oder es wird der Referenzverlauf eines Referenznetzwerkforums, d.h. der Referenzverlauf, welcher aus dem Referenznetzwerkforum entstammt, also aus diesem ermittelt worden ist, herangezogen.

[0015] Insbesondere können zur Vereitelung der Erkennung solche Referenzgrößen derart herangezogen werden, dass bei einer maschinengenerierten Eingabe von Texten Mittelwert und Standardabweichung der oben beschriebenen zeitlichen Größen an Mittelwert und Standardabweichung der Eingaben von Texten menschlicher Nutzer angepasst werden. Beispielsweise werden mittels eines Zufallsgenerators Abweichungen vom Mittelwert innerhalb der jeweiligen Standardabweichung erzeugt. Neben solchen expliziten Algorithmen kann auch für einzelne zeitliche Verläufe ein neuronales Netz verwendet werden. Auf diese Weise lässt sich das menschliche Eingabeverhalten nahezu exakt nachahmen.

[0016] Nachfolgend wird die Erfindung anhand eines in der Zeichnung dargestellten Ausführungsbeispiels näher erläutert:

[0017] Die einzige Zeichnung **Fig. 1** zeigt charakteristische Kenngrößen bei der Eingabe eines Textes schematisch in einer Prinzipskizze. Diese Kenngrößen werden zur erfindungsgemäßen Erkennung, ob der Text maschinengeneriert ist, herangezogen

oder aber werden erfindungsgemäß herangezogen, die Erkennung einer Maschinengenerierung dieses Textes zu vereiteln.

[0018] Der in **Fig. 1** gezeigte Text **T** ist der Text eines Netzwerkforums, beispielsweise eines Internetforums, und umfasst einzelne Wörter **W**. Die Wörter **W** sind aus einzelnen Zeichen und Buchstaben **A**, **B**, **C** aufgebaut, von denen einige Buchstaben **V**, **E**, **R**, Silben **S** bilden. Die einzelnen Zeichen und Buchstaben **A**, **B**, **C** werden als Zeichenstrom eingegeben, welcher in **Fig. 1** als horizontale zeitliche Aufeinanderfolge von Zeichen und Buchstaben **A**, **B**, **C** repräsentiert ist.

[0019] Erfindungsgemäß heranziehbar sind etwa die Gesamtdauer für die Eingabe des Textes **T**, welche sich durch die Differenz **DT** des Startzeitpunkts **ST** und des Endzeitpunkts **ET** der Eingabe des Textes **T** bestimmen lässt. Beispielsweise lässt sich der Startzeitpunkt **ST** durch den Aufruf eines Editors und der Endzeitpunkt **ET** durch den Abschluss einer Texteingabe, etwa mittels der Eingabetaste oder mittels des Schließens des Editors, erfassen.

[0020] Ferner lassen sich Zeitdauern **T1** für die aktuelle oder die durchschnittliche Eingabe eines Zeichens oder Buchstabens **A**, **B**, **C** sowie die Zeitdauer **T2** für die aktuelle oder durchschnittliche Eingabe eines Wortes **W**, das heißt die Zeitdauer zwischen einem Startzeitpunkt eines ersten Wortes **SW1** und den Startzeitpunkt für die Eingabe eines zweiten Wortes **SW2**, ermitteln.

[0021] Solche Zeitdauern lassen sich grundsätzlich rechnerisch durch Division der Gesamtzeitdauer **DT** des Textes durch die Länge des Textes in Wörtern **W** und/oder Zeichen und/oder Buchstaben **A**, **B**, **C** ermitteln. Im dargestellten Ausführungsbeispiel aber werden diese Zeitdauern **T2** der Eingabe eines Wortes **W** und/oder der Zeitdauern **T1** eines Zeichens und/oder Buchstabens **A**, **B**, **C** mittels eines nutzerseitig installierten Programms, beispielsweise eines mittels eines Webbrowsers gestarteten Scripts, erfasst, so wie es von der eingabesynchronen Autovervollständigung von Eingaben in Eingabemasken von Suchmaschinen bekannt ist.

[0022] Diese zeitlichen Größen können erfasst und ausgewertet werden, so dass eine statistische Verteilung der Zeitdauern für die Gesamtzeitdauer **DT** des Textes oder die einzelnen Zeitdauern **T1** für die Eingabe eines Zeichens und/oder Buchstabens **A**, **B**, **C** oder **T2** für die Eingabe eines Wortes **W** vorliegen. Diese Zeitdauern **DT**, **T1**, **T2** und/oder ihre statistischen Verteilungen können mit Referenzeingaben menschlicher Nutzer verglichen werden, so dass Zeitdauern **DT**, **T1**, **T2** oder statistische Verteilungen dieser Zeitdauern **DT**, **T1**, **T2**, welche mit einem Mindestmaß von typischen Zeitdauern **DT**, **T1**, **T2** und/oder ihrer jeweiligen statistischen Verteilungen bei Texten von menschlichen Nutzern abweichen, auf eine Maschinengenerierung des Textes **T** hindeuten.

[0023] Umgekehrt können solche statistischen Verteilungen auch mittels eines Referenznetzwerkforums mit einem oder mehreren Nutzern gewonnen und nachfolgend bei der Maschinengenerierung von Texten genutzt werden, indem diese Verteilungen an Verteilungen menschlicher Nutzer angepasst werden, so dass die Erkennung eines maschinengenerierten Textes **T** vereitelt wird.

Patentansprüche

1. Verfahren zur Erkennung eines maschinengenerierten Textes (T) in einem Netzwerkforum, bei welchem der zeitliche Verlauf (DT, T1, T2) einer Eingabe des Textes (T) zur Erkennung herangezogen wird.

2. Verfahren zur Vereitelung einer Erkennung eines maschinengenerierten Textes (T) in einem Netzwerkforum, bei welchem der zeitliche Verlauf (DT, T1, T2) einer Eingabe des Textes (T) zur Vereitelung der Erkennung herangezogen wird.

3. Verfahren nach einem der vorhergehenden Ansprüche, bei welchem der zeitliche Verlauf (DT, T1, T2) der Eingabe des Textes (T) zumindest die Gesamtzeitdauer (DT) der Eingabe des Textes (T) umfasst.

4. Verfahren nach einem der vorhergehenden Ansprüche, bei welchem der zeitliche Verlauf (DT, T1, T2) zumindest die Zeitdauer (T2) der Eingabe zumindest eines Wortes (W) und/oder die Zeitdauer (T1) zumindest eines Zeichens und/oder Buchstabens (A, B, C) des Textes (T) umfasst.

5. Verfahren nach einem der vorhergehenden Ansprüche, bei welchem der zeitliche Verlauf (DT, T1, T2) die Zeitdauer der Eingabe zumindest einer Silbe (S) umfasst.

6. Verfahren nach einem der vorhergehenden Ansprüche, bei welchem der zeitliche Verlauf (DT, T1, T2) ein Maß für die Streuung der Gesamtzeitdauer (DT) oder einer Zeitdauer (T1, T2) gemäß einem der Ansprüche 4 oder 5 umfasst.

7. Verfahren nach einem vorhergehenden Ansprüche, insbesondere nach einem der der Ansprüche 4 bis 6, bei welchem bei dem zeitlichen Verlauf ein Anteil kopierten Textes berücksichtigt wird.

8. Verfahren nach einem der vorhergehenden Ansprüche, bei welchem der zeitliche Verlauf den zeitlichen Verlauf der Eingabe von nebeneinanderliegenden Zeichen und/oder Buchstaben (A, B, C) und/oder nicht mit derselben Hand zu betätigenden

Zeichen und/oder Buchstaben auf einer Normtastatur, insbesondere einer QWERTY-Tastatur oder einer QWERTZ-Tastatur oder einer Dvorak-Tastatur, umfasst.

9. Verfahren nach einem der vorhergehenden Ansprüche, bei welchem der zeitliche Verlauf mit einem Referenzverlauf verglichen wird.

10. Verfahren nach dem vorhergehenden Anspruch, bei welchem der Referenzverlauf mittels eines Referenznetzwerkforums ermittelt oder der Referenzverlauf aus einem Referenznetzwerkforum herangezogen wird.

Es folgt eine Seite Zeichnungen

Anhängende Zeichnungen

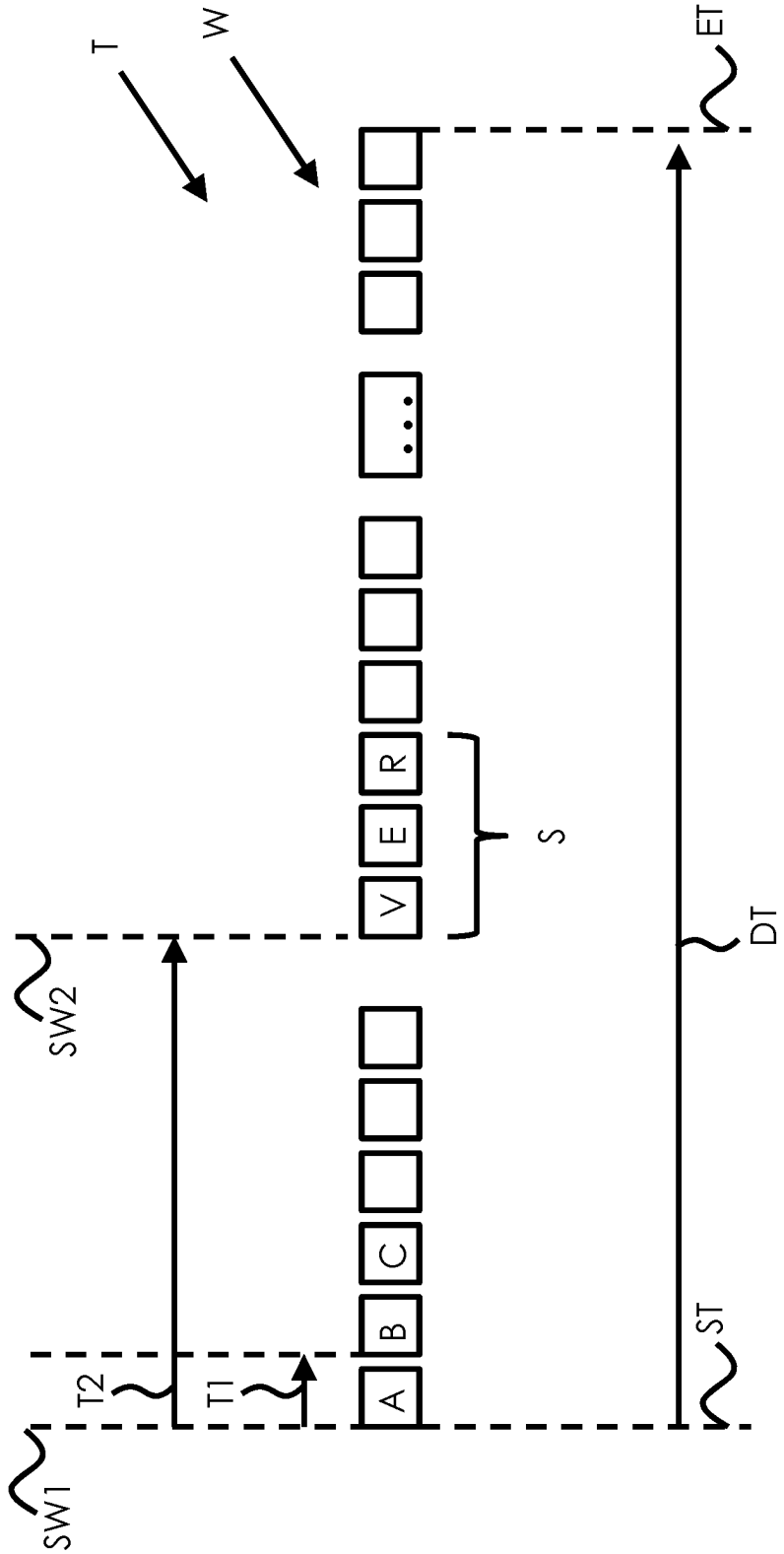


Fig. 1