(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2019-193129
(P2019-193129A)
(43)公開日　令和1年10月31日(2019.10.31)

(51)Int.Cl.			F I			テーマコード（参考）
H03M	7/42	(2006.01)	H03M	7/42		5J064
H04L	12/721	(2013.01)	H04L	12/721	Z	5K030

審査請求 未請求　請求項の数 10　ＯＬ　（全 23 頁）

(21)出願番号　特願2018-84774(P2018-84774)
(22)出願日　平成30年4月26日(2018.4.26)

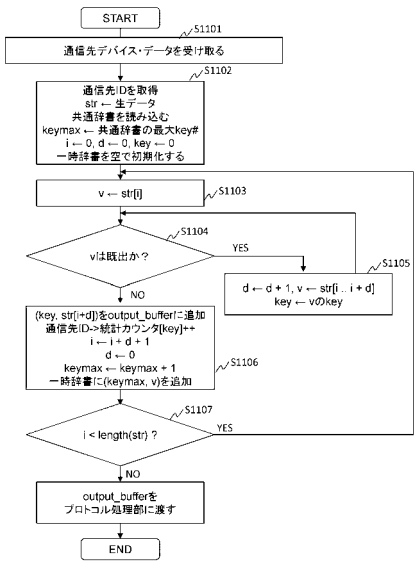
(71)出願人　000005108
株式会社日立製作所
東京都千代田区丸の内一丁目６番６号
(74)代理人　110000062
特許業務法人第一国際特許事務所
(72)発明者　谷田　直輝
東京都千代田区丸の内一丁目６番６号　株式会社日立製作所内
(72)発明者　デーンホフ　イェンス
東京都千代田区丸の内一丁目６番６号　株式会社日立製作所内
(72)発明者　對馬　雄次
東京都千代田区丸の内一丁目６番６号　株式会社日立製作所内

最終頁に続く

(54)【発明の名称】データ転送装置及びデータ転送方法、並びにデータ転送装置を備えたネットワークシステム

(57)【要約】
【課題】通信内容の変化に対応可能であり、高圧縮通信することができる技術を提供する。
【解決手段】共通辞書学習部、符号化部、復号化部、を備えたデータ通信装置において、事前に学習期間における通信データから圧縮用の共通辞書を作成し、通信データのデータ転送時に共通辞書に一時的に作成した一時辞書を付加してワンパス（ｏｎｅ－ｐａｓｓ）によるデータ転送する。また、共通辞書を作成する上での共通辞書学習の際、類似する通信先デバイスの学習用データを纏める。
【選択図】　図１１

図11



【特許請求の範囲】
【請求項１】
通信デバイスの学習期間における通信データから圧縮用の共通辞書を構築する共通辞書学習部と、
ネットワークへ送信するデータをもとに前記共通辞書へエントリを追加しながら前記送信するデータの圧縮を行う符号化部と、
前記ネットワークを介して受信したデータに含まれる一時辞書と前記圧縮されたデータをもとに前記受信したデータの復号化を行う復号化部と、
前記共通辞書へ追加したエントリを一時辞書として、前記圧縮したデータと共に前記ネットワークへ送信し、前記ネットワークを介して受信した前記一時辞書を含む前記圧縮されたデータを受信する転送制御部と、
を備えることを特徴とするデータ転送装置。
【請求項２】
請求項１に記載されたデータ転送装置において、
学習データ選別部をさらに有し、
前記学習データ選別部は、
通信先デバイスごとに、通信パラメータの統計を学習デバイスリスト保存部に記憶し、当該統計した統計値の差が所定の値以下である通信先デバイスが複数ある場合、前記通信先デバイスのうち一部の通信データを学習用の通信データから除外する
ことを特徴とするデータ転送装置。
【請求項３】
複数の通信先デバイス間で通信データを、ネットワークを介して転送するデータ転送装置であって、
共通辞書学習部、符号化部、復号化部、転送制御部、を有し、
前記共通辞書学習部は、
前記転送制御部を介して受信した学習期間における通信データのコピーから前記通信データを圧縮するための情報を含む共通辞書を構築し、
前記符号化部は、
前記転送制御部から通信先デバイスのデバイス情報及び通信データを受け、当該通信データをもとに前記共通辞書へエントリを追加し、前記通信先デバイスから送信する前記通信データの圧縮を行い、
前記復号化部は、
前記符号化部にて圧縮された通信データを受信し、当該圧縮された通信データに含まれる一時辞書をもとに当該圧縮されたデータの復号化を行い、
前記転送制御部は、
前記共通辞書に追加したエントリを一時辞書として、前記符号化部にて符号化された圧縮データとともに、前記ネットワークを介して前記通信先デバイスへ送信し、
また、前記復号化部にて復号化された非圧縮データを、前記ネットワークを介して受信した前記一時辞書を含む前記圧縮データを受信する
ことを特徴とするデータ転送装置。
【請求項４】
請求項１又は３に記載されたデータ転送装置において、
前記共通辞書は、前記通信データを圧縮するための情報を格納するエントリを含み、当該エントリは、符号、記号列を含む、前記通信データを圧縮するための情報であり、
前記符号化部は、前記通信データを符号化するとき、当該通信データに記号列が含まれているとき、前記記号列を前記符号に変換し、
前記復号化部は、前記圧縮されたデータを復号化するとき、前記符号を前記記号列に復元する
ことを特徴とすることを特徴とするデータ転送装置。
【請求項５】

請求項１又は３に記載されたデータ転送装置において、
前記共通辞書学習部は、
前記共通辞書を構築するとき、
通信先デバイスごとに通信パラメータの統計を記憶し、
前記統計の値が所定の値以下の場合、
前記通信先デバイスのうち一部の通信データを学習用データから除外する
ことを特徴とするデータ転送装置。
【請求項６】
第１ネットワークと第２ネットワークの間に請求項１又は請求項３又は請求項６に記載されたデータ転送装置を設け、複数の端末の間で通信データの転送を可能とするネットワークシステム。
【請求項７】
通信デバイスの学習期間における通信データから圧縮用の共通辞書を構築する共通辞書学習ステップと、
ネットワークへ送信するデータをもとに前記共通辞書へエントリを追加しながら前記送信するデータの圧縮を行う符号化ステップと、
前記ネットワークを介して受信したデータに含まれる一時辞書と前記圧縮されたデータをもとに前記受信したデータの復号化を行う復号化ステップと、
前記共通辞書へ追加したエントリを一時辞書として、前記圧縮したデータと共に前記ネットワークへ送信し、前記ネットワークを介して受信した前記一時辞書を含む前記圧縮されたデータを受信する転送制御ステップと、
を備えることを特徴とするデータ転送方法。
【請求項８】
複数の通信先デバイス間で通信データを転送するデータ転送方法であって、
転送制御部を介して受信した学習期間における通信データのコピーから前記通信データを圧縮するための情報を含む共通辞書を構築する共通辞書学習ステップと、
前記転送制御部から通信先デバイスのデバイス情報及び通信データを受け、当該通信データをもとに前記共通辞書へエントリを追加し、前記通信先デバイスから送信する前記通信データを圧縮する符号化ステップと、
前記符号化ステップにて圧縮された通信データを受信し、当該圧縮された通信データに含まれる一時辞書をもとに当該圧縮されたデータを復号化する復号化ステップと、
前記共通辞書に追加したエントリを一時辞書として、前記符号化部にて符号化された圧縮データとともに、前記通信先デバイスに送信し、前記復号化部にて復号化された非圧縮データを、前記ネットワークを介して受信した前記一時辞書を含む前記圧縮データを受信する転送制御ステップと、を有する
ことを特徴とするデータ転送方法。
【請求項９】
請求項７又は８に記載されたデータ転送方法において、
前記共通辞書は、前記通信データを圧縮するための情報を格納するエントリを含み、当該エントリは、符号、記号列からなり、
前記符号化ステップは、前記通信データを符号化するとき、当該通信データに記号列が含まれているとき、前記記号列を前記符号に変換し、
前記復号化ステップは、前記圧縮された通信データを復号化するとき、前記符号を前記記号列に復元する
ことを特徴とすることを特徴とするデータ転送方法。
【請求項１０】
請求項７又は８に記載されたデータ転送方法において、
前記共通辞書学習ステップは、
前記共通辞書を構築するとき、
前記通信先デバイスごとに通信パラメータの統計を記憶するステップ、

前記統計の値が所定の値以下の場合、前記通信先デバイスのうち一部の通信データを学習用データから除外するステップを含む
ことを特徴とするデータ転送方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、データ転送装置及びデータ転送方法、並びにデータ転送装置を備えたネットワークシステムに関する。
【背景技術】
【０００２】
様々な設備や機器を広域ネットワークに接続し、運用・監視・保守する「モノ」のネットワーク化、所謂、ＩｏＴ（Ｉｎｔｅｒｎｅｔ　ｏｆ　Ｔｈｉｎｇｓ）が普及しつつある。
ＩｏＴでは、ログ情報等の類似データを繰り返し、通信で収集するという特性がある。ＩｏＴにおける通信では、例えば、３Ｇ／ＬＴＥ／ＬＰＷＡ（Ｌｏｗ　Ｐｏｗｅｒ，Ｗｉｄｅ　Ａｒｅａ）等の広域無線通信等が広く利用される。広域無線通信は、大量の通信を行うためには通信コストが高くなるといという課題がある。
【０００３】
斯様な課題を是正する本技術分野の背景技術の一つとして、特開２０１７－１７３９８２号公報（特許文献１）に記載の技術が提案されている。この公報には、「ログ収集装置４００は、ログ生成装置からテキストログを受信するログ受信部４１０と、受信されたテキストログに基づいて、テキスト圧縮を行うための圧縮辞書を生成する辞書生成部４４０と、ログ生成装置に対し、生成された圧縮辞書を送信して、当該圧縮辞書の送信後に送信されるテキストログに対して当該圧縮辞書を使用した圧縮処理を行うことを指示する辞書送信部４６０と、当該圧縮辞書の送信後に受信されたテキストログに対して圧縮辞書を使用して伸張処理を行う伸張処理部４２０と、を有する。」という記載がある（要約書参照）。
【先行技術文献】
【特許文献】
【０００４】
【特許文献１】特開２０１７－１７３９８２号公報
【発明の概要】
【発明が解決しようとする課題】
【０００５】
特許文献１では、圧縮辞書を使用して伸張処理を行うことにより、ログ生成装置の処理負荷の増大を抑えた状態で、低負荷なテキストログ伝送を可能とするという効果を奏する。
しかし、特許文献１には、辞書を更新する手法、また、辞書の生成に利用するデータの選別については明示なく、何ら考慮されていない。そのため、ログ生成装置の処理負荷の増大を抑えた状態で、低負荷なテキストログ伝送を可能とするという効果に留まる。
つまり、特許文献１のように、事前に生成した辞書を用いてデータを圧縮する通信圧縮方式のシステムでは、システムの更新によって事前に想定していた通信内容と実際の通信内容に差異が生じた場合、圧縮率が想定より低下するという課題があった。
【０００６】
また、実際の通信内容を元に辞書を生成する場合、大規模なＩｏＴシステム等において全てのデータをもとに辞書を生成すると、大量の計算機リソースが必要となる課題があった。
【０００７】
そこで、本発明では、上述した課題に鑑み、通信内容の変化に対応可能であり、高圧縮通信することができる技術を提供することを目的とする。

【課題を解決するための手段】
【０００８】
上記課題を解決するために、代表的な本発明のデータ転送装置及びネットワークシステム、並びにデータ転送方法の一つは、事前に学習期間における通信データから圧縮用の共通辞書を作成し、前記通信データのデータ転送時に前記共通辞書に一時的に作成した一時辞書を付加してワンパス（ｏｎｅ－ｐａｓｓ）によるデータ転送するものであって、前記共通辞書により、前記通信データにおけるパケット間の重複を排除し、前記一時辞書により、前記通信データのパケットごとの差異を吸収し、システムの更新によって前記通信データに変化が生じた場合における通信効率の低下を抑制し、高い圧縮率と細かな前記通信データの変化への対応を実現するものである。
【０００９】
また、前記共通辞書を作成する上での共通辞書学習の際、類似する通信先デバイスの学習用データを纏め、学習量を低下させ、負荷を軽減するものである。
【発明の効果】
【００１０】
本発明によれば、データ転送において、通信内容が変化しても高効率に通信圧縮を行うことができ、通信コストの削減や通信速度の向上を実現することができる。
上記した以外の課題、構成及び効果は、以下の実施例の説明によって明らかにされる。
【図面の簡単な説明】
【００１１】
【図１】本発明のデータ転送装置をネットワークシステムに適用した一構成例を示す図。
【図２】本発明のデータ転送装置のハードウェア構成及び主記憶装置におけるソフトウェア構成の一例を示す図。
【図３】本発明のデータ転送装置のソフトウェア構成の論理的な接続関係を示す図である。
【図４】本発明のデータ転送装置がＬＡＮを介して送受信するＴＣＰパケットのフォーマットの一例を示す図。
【図５】本発明のフロー情報の一例を示す図。
【図６】本発明の学習デバイスリストの一例を示す図。
【図７（Ａ）】本発明の共通辞書の一例を示す図。
【図７（Ｂ）】本発明の一時辞書の一例を示す図。
【図８】本発明のデータ転送処理の流れを説明するシーケンス図。
【図９】本発明のデータ転送処理の流れを説明するシーケンス図。
【図１０】本発明の転送制御部が実行する処理を説明するフローチャート。
【図１１】本発明の符号化部が実行する処理を説明するフローチャート。
【図１２】本発明の復号化部が実行する処理を説明するフローチャート。
【図１３】本発明の共通辞書学習部が実行する処理を説明するフローチャート。
【図１４】本発明の共通辞書学習部において、共通辞書を作成する処理の詳細を説明するフローチャート。
【図１５】本発明のデータ転送装置をネットワークシステムに適用した他の構成例を示す図。
【図１６】本発明の学習データ選別部が実行する処理を説明するフローチャート。
【図１７】図１６の学習データ選別部の類似デバイス登録処理済判定処理の詳細処理手順を示すフローチャート。
【図１８】通信先デバイスのデバイス情報の一構成例を示す図。
【発明を実施するための形態】
【００１２】
本実施形態では、基本的に、通信データを圧縮する機能が、通信データを中継、転送する装置に実装される場合について説明する。また、データ転送として、ＴＣＰ／ＩＰ通信におけるデータ転送の例を挙げて説明する。

ただし、本発明は、ＴＣＰ／ＩＰにおけるデータ転送に限定されない。

【００１３】

本発明は、データ転送時に共通辞書に一時辞書を付加し、ワンパス（ｏｎｅ－Ｐａｓｓ）による圧縮転送を可能とする。

そして、共通辞書により、通信データのパケット間の重複を排除し、一時辞書により、パケットごとの差異を吸収することで、高い圧縮率と細かな通信内容の変化への対応を実現するものである。

【００１４】

また、ＩｏＴにおいて、類似デバイスによる類似通信が多いことに着目し、共通辞書を学習する際、類似デバイスの学習データを纏めることで学習量を低下させ、負荷を軽減するものである。

【００１５】

以下、その実施例について図面を参照して説明する。なお、各図において、共通する構成については同一の参照符号が付されている。

【実施例１】

【００１６】

図１は、本発明のデータ転送装置をネットワークシステムに適用した一構成例を示す図である。

【００１７】

ネットワークシステムは、端末１００、データ転送装置１１０、ネットワーク１４０、から構成される。

【００１８】

本例では、端末１００は、複数の端末１００－１、１００－２を含み、データ転送装置１１０は、複数のデータ転送装置１１０－１、１１０－２、を含む。

【００１９】

ネットワーク１４０は、複数の端末１００－１、１００－２、複数のデータ転送装置１１０－１、１１０－２を接続する、例えば、ＬＡＮ（Ｌｏｃａｌ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）、ＷＡＮ（Ｗｉｄｅ　Ａｒｅａ　Ｎｅｔｗｏｒｋ）、等からなり、複数の中継装置１３０－１、１３０－２、１３０－３、を構成する。

なお、本発明は、ネットワークシステムに含まれる各装置及びネットワークの数には限定されない。

【００２０】

すなわち、端末１００－１とデータ転送装置１１０－１及び端末１００－２とデータ転送装置１１０－２は、ＬＡＮ１４０－１の中継装置１３０－１及びＬＡＮ１４０－２の中継装置１３０－３を介して互いに接続される。また、複数のデータ転送装置１１０－１とデータ転送装置１１０－２は、ＷＡＮ１２０の中継装置１３０－２を介して互いに接続される。

【００２１】

各ネットワークの接続方式は、無線及び有線のいずれもよい。ＬＡＮ１４０及びＷＡＮ１２０には、中継装置１３０が一つ以上含まれる。

また、ネットワークは、各装置を接続するネットワークの種別に限定されず、専用回線及びインターネット等のネットワークでもよい。また、端末１００及びデータ転送装置１１０は、直接接続されてもよい。複数のデータ転送装置１１０は、直接接続されてもよい。

【００２２】

端末１００は、データ通信機能有し、通信先通信デバイスや通信元通信デバイスとなる装置であって、例えば、物理計算機及びプロキシ装置、等が考えられる。また、仮想計算機を用いて実現してもよい。

端末１００は、図示していないが、データ通信機能を実現する演算装置、主記憶装置、及び有するネットワークインターフェースを有する。主記憶装置は、データ通信機能を実

行するプログラムやデバイス情報、データ、等を記憶する。
【００２３】
中継装置１３０は、ＬＡＮ１４０及びＷＡＮ１２０等のネットワークを構成する装置であり、スイッチ、ルータ、及びゲートウェイ装置、等が考えられる。
また、中継装置１３０は、図示していないが、データの通信機能を実現する演算装置、主記憶装置、及びネットワークインターフェースを有する。
【００２４】
データ転送装置１１０は、端末１００（１００－１，１００－２）間のＴＣＰ通信の終端及び中継を行う装置であり、端末１００間で送受信されるデータを転送する。
【００２５】
本例のデータ転送装置１１０は、フロー単位で通信を制御する機能を有するものとする。
例えば、データ転送装置１１０－１は、端末１００－１と端末１００－２との間のＴＣＰ通信１５０において、ＴＣＰ通信１５０を終端し、ＬＡＮ１４０－１を介して端末１００－１との間でＴＣＰ通信１５１を行う。
【００２６】
なお、ＴＣＰ通信１５１では、データ転送装置１１０－１は、ＴＣＰ通信１５０において、端末１００－２が用いるＩＰアドレス及びポート番号を、データ転送装置１１０－１のＩＰアドレス及びポート番号として用いる。
【００２７】
同様に、データ転送装置１１０－２は、ＴＣＰ通信１５０において、ＴＣＰ通信１５０を終端し、ＬＡＮ１４０－２を介して端末１００－２とＴＣＰ通信１５３を行う。
データ転送装置１１０－２は、ＴＣＰ通信１５３では、ＴＣＰ通信１５０において端末１００－１が用いるＩＰアドレス及びポート番号を、データ転送装置１１０－２のＩＰアドレス及びポート番号として用いる。
【００２８】
また、データ転送装置１１０－１及びデータ転送装置１１０－２は、ＷＡＮ１２０を介して、ＴＣＰ通信１５２を行う。
【００２９】
ＴＣＰ通信１５２では、データ転送装置１１０－１は、端末１００－１のＩＰアドレス及びポート番号を用い、また、データ転送装置１１０－２は、端末１００－２のＩＰアドレス及びポート番号を用いる。
【００３０】
上述したＴＣＰ通信によって、端末１００－１及び端末１００－２は、端末１００間で直接通信するＴＣＰ通信１５０を行っている場合と同様の動作を行う。
【００３１】
また、本例のデータ転送装置１１０は、転送するデータを圧縮・展開して帯域を節約するように通信を制御する。
具体的には、データ転送装置１１０（１１０－１）は、転送元（通信先）の端末１００（１００－１）からＬＡＮ１４０（１４０－１）を経由して受信したデータを圧縮し、ＷＡＮ１２０へ転送する。
【００３２】
データ転送装置１１０（１１０－２）は、ＷＡＮ１２０から受信したデータを展開し、ＬＡＮ１４０（１４０－２）を経由して転送先（通信先）の端末１００（１００－２）にデータを送信する。
【００３３】
上述したようなデータの転送制御を行うことによって、ＷＡＮ１２０へのデータ転送量が削減される。その結果として、通信コストの削減及び端末１００間の通信速度の高速化を実現できる。
【００３４】

図２は、本発明のデータ転送装置１１０のハードウェア構成及びその一部であるハードウェア（主記憶装置２０２）のソフトウェア構成の一例を示す図であり、図３は、データ転送装置１１０のソフトウェア（主記憶装置２０２）の論理的な接続関係を示す図である。

【００３５】

同図において、データ転送装置１１０は、演算装置２０１、主記憶装置２０２、二次記憶装置２０３、及び複数のネットワークインターフェース２０４を有する。

【００３６】

本例では、各装置、各インターフェースは、それぞれ１つであるが、二つ以上あってもよい。また、仮想的なハードウェアを用いて実現してもよい。

例えば、一つのネットワークインターフェース２０４上に複数の論理的なネットワークインターフェースを構成し、論理的なネットワークインターフェースを用いることもできる。

【００３７】

データ転送装置１１０における演算装置２０１、主記憶装置２０２、二次記憶装置２０３、ネットワークインターフェース２０４は、システムバス２０５を介して互いに接続され、システムバス２０５を介して互いに通信を行う。

【００３８】

なお、データ転送装置１１０における各ハードウェアの数は二つ以上でもよい。また、各ハードウェアは、複数のシステムバス２０５を介して、互いに接続されてもよく、また、それぞれ直接接続されてもよい。

【００３９】

データ転送装置１１０は、図示していないが、入出力装置を有してもよい。入出力装置には、キーボード、マウス、タッチパネル、及びディスプレイ等が含まれる。

【００４０】

演算装置２０１は、主記憶装置２０２に格納される各機能部、つまり、共通辞書学習機能を実行する共通辞書学習部（共通辞書学習プログラム）２１１、符号化機能を実行する符号化部（符号化プログラム）２１２、復号化機能を実行する復号化部（復号化プログラム）２１３、転送制御機能を実行する転送制御部（転送制御プログラム）２１４、プロトコル処理を実行するプロトコル処理部（プロトコル処理プログラム）２１５、等を実行する装置である。

【００４１】

演算装置２０１は、例えば、ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）及びＧＰＵ（Ｇｒａｐｈｉｃｓ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）等が考えられる。

演算装置２０１が上述した各機能部（プログラム）を実行することによってデータ転送装置１１０が有する機能を実現できる。

【００４２】

以下、各機能部（モジュール）を主体にした処理について説明するが、演算装置２０１において、当該機能部を実現するプログラムを実行していることを示す。

【００４３】

主記憶装置２０２は、演算装置２０１が実行するプログラム及び当該プログラムによって使用される情報を格納する保存部、例えば、フロー情報保存部２１６、学習デバイスリスト保存部２１７、学習用データ保存部２１８、共通辞書保存部２１９、等を有する。フロー情報、学習デバイスリスト、共通辞書、等については後述する。

【００４４】

主記憶装置２０２は、また、図示していないが、プログラムが使用するワークエリア及びバッファ等の一時記憶領域を含む。主記憶装置２０２としては、メモリ等が考えられる。

【００４５】

二次記憶装置２０３は、永続的にデータを格納する装置である。また、二次記憶装置２０３は、主記憶装置２０２に格納できないデータや後述する一時辞書７００’、等を一時的に格納する一時記憶領域を含む。

【００４６】

二次記憶装置２０３としては、例えば、ＨＤＤ（Ｈａｒｄ　Ｄｉｓｋ　Ｄｒｉｖｅ）及びＳＳＤ（Ｓｏｌｉｄ　Ｓｔａｔｅ　Ｄｒｉｖｅ）等が考えられる。

【００４７】

主記憶装置２０２に格納される各プログラム及び各情報は、二次記憶装置２０３に格納されてもよい。

この場合、演算装置２０１は、二次記憶装置２０３から、必要なプログラム及び情報を読み出し、主記憶装置２０２にプログラム及び情報をロードする。

【００４８】

ネットワークインターフェース２０４（２０４－１、２０４－２）は、ネットワーク（ＬＡＮ、ＷＡＮ）と接続するためのインターフェースである。

【００４９】

本例では、データ転送装置１１０は、ネットワークインターフェース２０４－１を介してＬＡＮ１４０（１４０－１，１４０－２）と接続し、また、ネットワークインターフェース２０４－２を介してＷＡＮ１２０と接続する。

【００５０】

次に、主記憶装置２０２及び二次記憶装置２０３に格納されるプログラム及び情報について説明する。

【００５１】

主記憶装置２０２は、データ転送装置１１０で実行するプログラムとして、共通辞書学習部（共通辞書学習プログラム）２１１、符号化部（符号化プログラム）２１２、復号化部（復号プログラム）２１３、転送制御部（転送制御プログラム）２１４、プロトコル処理部（プロトコル処理プログラム）２１５を格納する。

【００５２】

また、主記憶装置２０２は、情報として、フロー情報（図５参照）をフロー情報保存部２１６に格納し、学習デバイスリスト（図６参照）を学習デバイスリスト保存部２１７に格納し、学習用データを学習用データ保存部２１８に格納し、共通辞書（図７参照）を共通辞書格納部２１９に格納する。

【００５３】

フロー情報５００は、データ転送におけるフローを管理するための情報であって、その詳細は、後述するが、図５に示すとおりである。

【００５４】

学習デバイスリスト６００は、共通辞書７００を生成するときに用いる通信データの対象とする通信先デバイス（端末１００）の情報であって、その詳細は、後述するが、図６に示すとおりである。

【００５５】

学習用データは、共通辞書７００を生成するときに用いる通信データをコピーしたものである。

【００５６】

共通辞書７００は、データ転送装置１１０によって圧縮される通信で共通して利用される圧縮用辞書であって、共通辞書学習部（共通辞書学習プログラム）２１１による学習期間中、または運用中に作成され、また、更新され、事前に送受信側に展開される。また、共通辞書７００は、通信データの圧縮時・展開時にマージされる。その詳細は、後述するが、図７（Ａ）に示すとおりである。

【００５７】

共通辞書学習部２１１は、端末１００－１、１００－２の通信先デバイスの通信データから学習用データを保存する学習用データ保存部２１８から学習用データ（生データ）を

取得し、当該学習用データをもとに共通辞書７００を作成し、共通辞書格納部２１９に保存する共通辞書学習機能を実行するものである。その詳細は、後述する。

【００５８】

符号化部２１２は、転送制御部２１４から通信先デバイスのデバイス情報（図１８参照）及び通信データを受け、当該通信データをもとに共通辞書へエントリを追加し、端末１００－１又は端末１００－２から送信する通信データを圧縮する符号化を行う符号化機能を実行するものである。共通辞書へ追加されたエントリは、一時辞書７００’（図７（Ｂ）参照）として、符号化部２１２にて圧縮されたデータとともに復号化部２１３へ出力される。一時辞書は、通信データの送信時、パケット毎に送信さら、また、使い捨てされる。その詳細は、後述する。

【００５９】

復号化部２１３は、符号化部２１２にて圧縮された通信データを受信し、当該圧縮された通信データに含まれる一時辞書７００’をもとに当該圧縮されたデータの復号化を行う復号化機能を実行するものである。その詳細は、後述する。

【００６０】

転送制御部２１４は、共通辞書に追加したエントリを一時辞書７００’として、符号化部にて符号化された圧縮データとともに、ネットワークインターフェース２０４－１を介してＷＡＮ１２０に送信し、また、ＷＡＮ１２０からネットワークインターフェース２０４－１を介して一時辞書７００’を含む圧縮データを受信する制御機能を実行するものである。また、復号化部にて復号化された非圧縮データを、ネットワークインターフェース２０４－２を介して端末１００－２又は端末１００－１に送信する制御機能を実行するものである。その詳細は、後述する。

【００６１】

プロトコル処理部２１５は、ネットワークインターフェース２０４との間で通信を行うプロトコル処理を実行するものである。

【００６２】

なお、上述した各機能部は、ソフトウェアとして実現されているが、こられの機能の一部又は全てを演算装置２０１及びネットワークインターフェース２０４等のハードウェアを用いて実現してもよい。

【００６３】

以上述べた本実施例によれば、データを通信する端末に変更を施すことなく、通信内容が変化しても高効率に通信圧縮を行うことができる。

次に、図４を用いて、ＴＣＰ通信に使用されるＴＣＰパケットについて説明する。

【００６４】

図４は、データ転送装置１１０がＬＡＮ１４０を介して送受信するＴＣＰパケット４００のフォーマットの一例を示す図である。

【００６５】

ＴＣＰパケット４００は、ＭＡＣヘッダ４１０、ＩＰヘッダ４２０、ＩＰオプションヘッダ４３０、ＴＣＰヘッダ４４０、ＴＣＰオプションヘッダ４５０、及びＴＣＰペイロード４６０を含む。

【００６６】

ＭＡＣヘッダ４１０は、ＤＭＡＣ４１１、ＳＭＡＣ４１２、ＴＰＩＤ４１３、ＴＣＩ４１４、及びＴｙｐｅ４１５を含む。

【００６７】

ＤＭＡＣ４１１は、宛先ＭＡＣアドレスを表し、ＳＭＡＣ４１２は、送信元ＭＡＣアドレスを表し、ＴＰＩＤ４１３は、タグ付きフレームであること及びタグの種類を表し、ＴＣＩ４１４は、タグの情報を表し、Ｔｙｐｅ４１５は、ＭＡＣフレームタイプを表す。

【００６８】

ＴＣＩ４１４は、さらに、ＰＣＰ４１６、ＣＦＩ４１７、及びＶＩＤ４１８を含む。ＰＣＰ４１６は優先度を表し、ＣＦＩ４１７は、ＭＡＣアドレスが正規フォーマットであ

るかを表し、ＶＩＤ４１８は、ＶＬＡＮのＩＤを表す。
【００６９】
なお、ＶＬＡＮが使用されていないネットワークの場合、ＴＰＩＤ４１３及びＴＣＩ４１４は存在しない。この場合、データ転送装置１１０は、ＶＩＤ４１８が「０」であるものとして処理を行う。
【００７０】
ＩＰヘッダ４２０は、ＩＰ　ｌｅｎｇｔｈ４２１、ｐｒｏｔｏｃｏｌ４２２、ＳＩＰ４２３、及びＤＩＰ４２４を含む。
【００７１】
ＩＰ　ｌｅｎｇｔｈ４２１は、ＭＡＣヘッダ４１０を除くパケット長を表し、ｐｒｏｔｏｃｏｌ４２２は、プロトコル番号を表し、ＳＩＰ４２３は、送信元ＩＰアドレスを表し、ＤＩＰ４２４は、宛先ＩＰアドレスを表す。
【００７２】
ＩＰオプションヘッダ４３０は、０又は複数のＩＰオプション情報４３１を含む。
【００７３】
ＴＣＰヘッダ４４０は、ｓｒｃ．ｐｏｒｔ４４１、ｄｓｔ．ｐｏｒｔ４４２、ＳＥＱ４４３、ＡＣＫ４４４、ｆｌａｇ４４５、ｔｃｐ　ｈｌｅｎ４４６、及びｗｉｎ＿ｓｉｚｅ４４７を含む。
【００７４】
ｓｒｃ．ｐｏｒｔ４４１は、送信元ポート番号を表し、ｄｓｔ．ｐｏｒｔ４４２は、宛先ポート番号を表し、ＳＥＱ４４３は、送信シーケンス番号を表し、ＡＣＫ４４４は、受信シーケンス番号を表し、ｆｌａｇ４４５は、ＴＣＰフラグ番号を表し、ｔｃｐ　ｈｌｅｎ４４６は、ＴＣＰのヘッダ長を表し、ｗｉｎ＿ｓｉｚｅ４４７は、対向装置へ通知する広告ウインドウサイズを表す。
【００７５】
ＴＣＰオプションヘッダ４５０は、０又は複数のオプションを含む。例えば、ｏｐｔｉｏｎ　ｋｉｎｄ４５１、ｏｐｔｉｏｎ　ｌｅｎｇｔｈ４５２、及びオプション情報４５３等のオプションが含まれる。
【００７６】
ｏｐｔｉｏｎ　ｋｉｎｄ４５１は、オプションの種別を表し、ｏｐｔｉｏｎ　ｌｅｎｇｔｈ４５２は、オプション長を表し、オプション情報４５３は、オプションの種類に応じた情報を表す。
【００７７】
次に、図５～図７を用いて、データ転送装置１１０が管理する情報等について説明する。図５は、フロー情報保存部２１６のフロー情報５００の一例を示す図である。
【００７８】
フロー情報５００は、図示のようにフローを管理する情報を格納するエントリを含む。エントリは、Ｆｌｏｗ－ＩＤ５０１、ＶＬＡＮ５０２、ＬＡＮ－ＩＰ５０３、ＷＡＮ－ＩＰ５０４、ＬＡＮ－ＰＯＲＴ５０５、ＷＡＮ－ＰＯＲＴ５０６から構成される。
【００７９】
Ｆｌｏｗ－ＩＤ５０１は、フローを一意に識別するための識別情報である。ＶＬＡＮ５０２は、フローが属するＶＬＡＮの識別情報である。ＬＡＮ－ＩＰ５０３及びＬＡＮ－ＰＯＲＴ５０５は、ＬＡＮ１４０を介して接続される端末１００のＩＰアドレス及びポート番号である。ＷＡＮ－ＩＰ５０４及びＷＡＮ－ＰＯＲＴ５０６は、ＷＡＮ１２０を介して接続される端末１００のＩＰアドレス及びポート番号である。
【００８０】
図６は、学習デバイスリスト保存部２１７の学習デバイスリスト６００一例を示す図である。
【００８１】
学習デバイスリスト６００は、共通辞書生成のために利用する通信の通信先情報を管理

する情報を格納するエントリを含む。エントリは、Ｄｅｖｉｃｅ－ＩＤ６０１、Ｄｅｖｉｃｅ－ＩＰ６０２から構成される。

【００８２】

Ｄｅｖｉｃｅ－ＩＤ６０１は、通信先の端末（通信先デバイス）を一意に識別するための識別情報である。Ｄｅｖｉｃｅ－ＩＰ６０２は、通信先の端末（通信先デバイス）のＩＰアドレスである。

【００８３】

学習デバイスリスト６００は、学習するデバイスのリストだけでなく、学習しないデバイスのリストも掲載する。学習するデバイスにはフラグを立てる。

【００８４】

図７（Ａ）及び（Ｂ）は、共通辞書格納部２１９の共通辞書７００及び一時辞書７００’の一例を示す図である。

【００８５】

共通辞書７００及び一時辞書７００’は、通信データを圧縮するための情報を格納するエントリを含む。

共通辞書７００のエントリは、符号ｋｅｙ７０１と対応する記号列Ｖａｌｕｅ７０２、統計カウンタ７０３から構成される。一時辞書７００’のエントリは、符号ｋｅｙ７０１’と対応する記号列Ｖａｌｕｅ７０２’から構成される。

【００８６】

データ転送装置１１０は、通信データを圧縮するとき、送信元の通信データの部分列として、記号列７０２が含まれていると、対応する符号ｋｅｙ７０１に変換する。

また、データ転送装置１１０は、圧縮された通信データを復号、展開するとき、符号ｋｅｙ７０１を、対応する記号列Ｖａｌｕｅ７０２に復元する。

共通辞書７００の作成方法及び利用方法の詳細については、後述する。

【００８７】

次に、図８及び図９を用いて、データ転送処理の一連の流れについて説明する。

本例では、端末１００－１から端末１００－２へデータを転送する場合のデータ転送処理について説明する。

なお、端末１００－２から端末１００－１へデータを転送する場合も同様の処理となる。

【００８８】

図８は、端末１００－１から端末１００－２へのデータ転送処理の流れを説明するシーケンス図である。なお、ＴＣＰ通信における確認応答等のＴＣＰレイヤの処理については省略している。

図８のシーケンスに基づく動作は、以下のとおりである。

【００８９】

端末１００－１は、データ転送装置１１０－１へデータを送信する（ステップＳ８０１）。

【００９０】

データ転送装置１１０－１は、ＬＡＮ１４０－１を介して、端末１００－１からデータを受信した場合、共通辞書学習部２１１にて、転送制御部２１４から受信したデータをコピーし、コピーしたデータを用いて共通辞書学習のための処理を行う（ステップＳ８０２）。

【００９１】

次に、データ転送装置１１０－１は、符号化部２１２にて、転送制御部２１４から受信したデータを符号化するための処理を行う（ステップ８０３）。

【００９２】

データ転送装置１１０－１は、符号化部２１２にて符号化した圧縮データを、ＷＡＮ１２０を介して、データ転送装置１１０－２へ送信する（ステップＳ８０４）。圧縮データは、一時辞書７００’の情報（符号Ｋｅｙ７０１’，記号列Ｖａｌｕｅ７０２’）を含む

。
【００９３】
データ転送装置１１０－２は、ＷＡＮ１２０を介して、データ転送装置１１０－１から圧縮データを受信した場合、復号化部２１３にて、圧縮データを復号化するための処理を行う（ステップＳ８０５）。
【００９４】
データ転送装置１１０－２は、復号化部２１３にて復号したデータを、ＬＡＮ１４０－２を介して、端末１００－２へ送信する（ステップＳ８０６）。
【００９５】
図９は、アップロードで学習するケースを示し、端末１００－２から端末１００－１へのデータ転送処理の流れを説明するシーケンス図である。図９のシーケンスに基づく動作は、以下のとおりである。
【００９６】
端末１００－２が端末１００－１へデータを送信する場合は、まず、端末１００－２は、データ転送装置１１０－２へ送信する（ステップＳ９０１）。
【００９７】
データ転送装置１１０－２は、ＬＡＮ１４０－２を介して、端末１００－２からデータを受信した場合、符号化部２１２にて、転送制御部２１４から受信したデータを符号化するための処理を行う（ステップＳ９０２）。
【００９８】
データ転送装置１１０－２は、符号化部２１２にて符号化した圧縮データを、ＷＡＮ１２０を介して、データ転送装置１１０－１へ送信する（ステップＳ９０３）。圧縮データは、一時辞書７００’の情報（符号Ｋｅｙ７０１’，記号列Ｖａｌｕｅ７０２’）を含む。
【００９９】
データ転送装置１１０－１は、ＷＡＮ１２０を介して、データ転送装置１１０－２から圧縮データを受信した場合、復号化部２１３にて、圧縮データを復号化するための処理を行う（ステップＳ９０４）。
【０１００】
データ転送装置１１０－１は、復号化部２１３にて復号したデータを共通辞書学習部２１１にて、コピーし、このコピーしたデータ（図１３の生データ）を用いて、共通辞書７００を作成する共通辞書学習のための処理を行う（ステップＳ９０５）。
【０１０１】
データ転送装置１１０－１は、復号化部２１３にて復号したデータを、ＬＡＮ１４０－１を介して、端末１００－１へ送信する（ステップＳ９０６）。
【０１０２】
図８及び図９では、送信側で学習、受信側で共通辞書学習のための処理を行う２つの例を示したが、これは計算能力の高い片側（データセンタ側）で学習ができることを示す。要するに、共通辞書学習処理は、通信が集約される側で行うとよい。
【０１０３】
次に、図１０～図１４、図１８を用いて、データ転送装置１１０の各機能部が実行する処理の詳細を説明する。
【０１０４】
図１０は、転送制御部２１４が実行する処理を説明するフローチャートである。図１０のフローチャートに基づく動作は以下のとおりである。
【０１０５】
転送制御部２１４は、プロトコル処理部２１５から、受信したネットワーク・通信先のデバイス情報及び通信データを受け取る（ステップＳ１００１）。
【０１０６】
デバイス情報１８０は、通信先のデバイスごとにトラフック（Traffic）,圧縮（Compre

ssion），Hit for nth keyの統計をとったものである。例えば、図１８に示すように、デバイスＩＤ１８０１、ＶＬＡＮ１８０２、ＩＰ１８０３、Ｔｒａｆｉｃ１８０４、Ｃｏｍｐｒｅｓｓｉｏｎ１８０５、Ｈｉｔ　ｆｏｒ　ｎｔｈ　ｋｅｙ１８０６、を含む。デバイス情報１８０は、通信先のデバイスで類似する類似デバイスは纏めるとよい。これにより、学習に用いるデータの総量を大幅に削減することができる。これについては後述する。

【０１０７】

Ｈｉｔ　ｆｏｒ　ｎｔｈ　ｋｅｙ１８０６とは、辞書（共通辞書／一時辞書）において、ｎ番目に頻度が高い符号Ｋｅｙのカウント数である。

【０１０８】

転送制御部２１４は、通信先デバイスが、図６の学習デバイスリスト６００に掲載されているかを確認する（ステップＳ１００２）。

【０１０９】

ステップＳ１００２で通信先デバイスが、学習デバイスリスト６００に掲載されている場合（ＹＥＳ）は、学習用データ保存部２１８に、通信データのコピーを保存する（ステップＳ１００３）。

【０１１０】

ステップＳ１００２で通信先デバイスが学習デバイスリスト６００に掲載されていない場合（ＮＯ）、及び、ステップＳ１００３の完了後、転送制御部２１４は、通信データを受信したネットワーク１４０がＬＡＮ１４０－１、１４－２であるかＷＡＮ１２０であるかを確認する（ステップＳ１００４）。

【０１１１】

ステップＳ１００４でネットワーク１４０がＬＡＮ１４０－１、１４－２であると確認された場合（ＹＥＳ）、転送制御部２１４は、符号化部２１２を呼び出し、受信した通信データを符号化部２１２に送信する（ステップＳ１００５）。

【０１１２】

ステップＳ１００４でネットワーク１４０がＷＡＮ１２０であると確認された場合（ＮＯ）、転送制御部２１４は、復号化部２１３を呼び出し、受信した通信データを復号化部２１３に送信する（ステップＳ１００６）。

【０１１３】

図１１は、実施例１の符号化部２１２が実行する処理を説明するフローチャートである。図１１のフローチャートに基づく動作は以下のとおりである。

【０１１４】

符号化部２１２は、転送制御部２１４から、通信先のデバイス情報１８０及び通信データを受け取る（ステップＳ１１０１）。

【０１１５】

符号化部２１２は、図６の学習デバイスリスト６００を参照して、通信先ＩＤ（Ｄｅｖｉｃｅ－ＩＤ６０１、Ｄｅｖｉｃｅ－ＩＰ６０２）を取得し、ポインタｓｔｒの示す場所に転送制御部２１４から受け取った通信データをコピーし、また、共通辞書保存部２１９から共通辞書７００を読み込み、共通辞書７００に含まれる符号ｋｅｙ７０１のうち最大の符号ｋｅｙ７０１を変数ｋｅｙｍａｘに代入し、変数ｉ，ｄ，ｋｅｙに０を代入し、図７（Ｂ）に示す一時辞書７００’を空で初期化する（ステップＳ１１０２）。

【０１１６】

符号化部２１２は、変数ｖに、変数ｓｔｒのｉ番目の文字を代入する（ステップＳ１１０３）。

【０１１７】

符号化部２１２は、図７（Ａ）に示す共通辞書７００、もしくは図７（Ｂ）に示す一時辞書７００’の記号列Ｖａｌｕｅ７０２に、変数ｖの値が含まれるか調べる（ステップＳ１１０４）。

【０１１８】

ステップＳ１１０４で、変数ｖの値が含まれる場合（ＹＥＳ）、符号化部２１２は、変

数ｄに１を加算し、変数ｖに変数ｓｔｒのｉ番目からｉ＋ｄ番目の文字からなる文字列を代入し、符号ｋｅｙ７０１に変数ｖの値に対応する共通辞書７００、または一時辞書７００’の符号ｋｅｙ７０１’の値を代入し、ステップＳ１１０４に戻る（ステップＳ１１０５）。

【０１１９】

ステップＳ１１０４で、変数ｖの値が含まれない場合（ＮＯ）、符号化部２１２は、符号ｋｅｙ及び、変数ｖの最後の文字のタプルを、ｏｕｔｐｕｔ＿ｂｕｆｆｅｒに追加し、通信先毎に符号ｋｅｙのヒット回数をカウントし、変数ｉにｉ＋ｄ＋１を代入し、変数ｄに０を代入し、一時辞書７００’に変数ｋｅｙｍａｘに１を加算し、一時辞書７００’に符号ｋｅｙ７０１’が変数ｋｅｙｍａｘで記号列Ｖａｌｕｅ７０２が変数ｖであるエントリを追加する（ステップＳ１１０６）。

【０１２０】

符号化部２１２は、変数ｉの値と変数ｓｔｒの文字列の長さｌｅｎｇｔｈ（ｓｔｒ）を比較する（ステップＳ１１０７）。

【０１２１】

ここで、変数ｉが変数ｓｔｒの文字列の長さｌｅｎｇｔｈ（ｓｔｒ）よりも小さい場合（ＹＥＳ）、ステップＳ１１０３に戻る。

変数ｉが変数ｓｔｒの文字列の長さよりも長い場合（ＮＯ）、符号化部２１２は、ｏｕｔｐｕｔ＿ｂｕｆｆｅｒをプロトコル処理部２１５に渡す（ステップＳ１１０８）。

【０１２２】

上述したフローチャートに基づく処理において、ステップＳ１１０２における「ｋｅｙｍａｘ←共通辞書の最大Ｋｅｙ＃」及び、ステップＳ１１０６における「通信先ID->統計カウンタ[ｋｅｙ]++」はユニークな点であり、ステップＳ１１０４にて、「ｖは既出か？」の処理において、「ｏｎｅ－ｐａｓｓ」で検索可能である点に優位性がある。

【０１２３】

図１２は、復号化部２１３が実行する処理を説明するフローチャートである。図１２のフローチャートに基づく動作は以下のとおりである。

【０１２４】

復号化部２１３は、転送制御部２１４から、通信先のデバイス情報及び通信データを受け取り、学習デバイスリスト６００を参照して、通信先ＩＤ（Ｄｅｖｉｃｅ－ＩＤ６０１、Ｄｅｖｉｃｅ－ＩＰ６０２）を取得し、共通辞書７００の内容を一時辞書ｔｍｐ＿ｄｉｃにし、共通辞書７００のエントリ数を変数ｋｅｙｍａｘに代入し、受け取った通信データをポインタｂｉｎの示す場所にコピーし、変数ｉに０を代入する（ステップＳ１２０１）。

【０１２５】

復号化部２１３は、ポインタｂｉｎが示すデータの先頭ｎｂｉｔを符号ｋｅｙ、続く８ｂｉｔを変数ｖに代入し、通信先毎に符号ｋｅｙのヒット回数をカウントし、ポインタｔｍｐの示す場所に一時辞書７００’の符号ｋｅｙ７０１’が符号ｋｅｙである記号列Ｖａｌｕｅ７０２と変数ｖを結合した文字列をコピーし、ポインタｔｍｐの示す文字列をｏｕｔｐｕｔ＿ｂｕｆｆｅｒの末尾に追加し、変数ｋｅｙｍａｘに１を加算し、一時辞書７００’に符号ｋｅｙ７０１’が変数ｋｅｙｍａｘで、記号列Ｖａｌｕｅ７０２がポインタｔｍｐの示す文字列であるエントリを追加し、変数ｉに変数ｎと８を加算する（ステップＳ１２０２）。変数ｎは、符号ｋｅｙのｂｉｔ幅である。

【０１２６】

復号化部２１３は、変数ｉとポインタｂｉｎが指し示すデータの長さＬｅｎｇｔｈ（ｂｉ）を比較する（ステップＳ１２０３）。

【０１２７】

ステップＳ１２０３で、変数ｉがデータの長さＬｅｎｇｔｈ（ｂｉ）よりも小さい場合（ＹＥＳ）、ステップＳ１２０２に戻る。

【０１２８】

ステップＳ１２０３で、変数ｉがデータの長さＬｅｎｇｔｈ（ｂｉ）よりも大きい場合（ＮＯ）、ｏｕｔｐｕｔ＿ｂｕｆｆｅｒを返す（ステップＳ１２０４）。

【０１２９】

図１３は、共通辞書学習部２１１が実行する処理を説明するフローチャートである。図１３のフローチャートに基づく動作は以下のとおりである。

【０１３０】

共通辞書学習部２１１は、転送制御部２１４がステップＳ１００３にて通信データのコピーを学習用データ保存部２１８に保存することを契機に実行される。

【０１３１】

共通辞書学習部２１１は、学習用データ保存部２１８の学習用データ（生データ）を取得する（ステップＳ１３０１）。

【０１３２】

共通辞書学習部２１１は、取得した学習用データ（生データ）が学習に十分であるか否かを判定する（ステップＳ１３０２）。

【０１３３】

取得した学習用データ（生データ）が学習に十分であるか否かの判定は、所定の日数以上、学習用データを取得したか否かで判定する。あるいは所定のバイト数以上の学習用データを取得したか否かで判定してもよいし、他の基準を設けてもよい。

【０１３４】

ステップＳ１３０２で、十分な学習用データがないと判定した場合（ＹＥＳ）、さらに学習用データを取得するため、共通辞書学習部２１１による共通辞書を作成する処理を終える。

【０１３５】

十分な学習用データがあると判定した場合（ＮＯ）は、共通辞書学習部２１１にて、取得した学習用データを用いて共通辞書を作成する（ステップＳ１３０３）。

共通辞書を作成する処理の詳細については、図１４を用いて後述する。

【０１３６】

共通辞書を作成した後は、さらに学習するか否かを判定する（ステップＳ１３０４）。

ステップＳ１３０４における判定には、所定の変数ｎのパケットに対して、符号化部２１２による符号化処理を行う際に、ステップＳ１１０２で読み込む共通辞書７００を、既存の共通辞書と新規学習した共通辞書の２つの場合で実行する。

そして、既存の共通辞書と新規に学習して作成した共通辞書の圧縮率を比較し、新規に学習して作成した共通辞書の圧縮率が既存の共通辞書の圧縮率よりも低い場合は、さらに学習すると判定し、新規に学習して作成した共通辞書が既存の共通辞書の圧縮率よりも高い場合は、共通辞書学習部２１１による学習処理を終えると判定する。

所定の変数ｎは、システムの通信量に合わせて変化させて良い。

【０１３７】

ステップＳ１３０４でさらに学習すると判定した場合は、共通辞書学習部２１１による学習処理を終える。

ステップＳ１３０４でさらに学習しないと判定した場合（ＹＥＳ）、作成した共通辞書を新規共通辞書として保存して処理を終える(ステップS１３０５)。

【０１３８】

図１４は、共通辞書学習部２１１における共通辞書を作成する処理のステップＳ１３０３の詳細を説明するフローチャートである。図１４のフローチャートに基づく動作は以下のとおりである。

【０１３９】

共通辞書学習部２１１は、共通辞書を初期化し、新規辞書とする。そして、ポインタｓｔｒの示す場所に学習用データをコピーし、０を変数ｋｅｙｍａｘに代入し、変数ｉ，ｄ，ｋｅｙに０を代入し、統計カウンタをクリアする（ステップＳ１４０１）。

【０１４０】

共通辞書学習部２１１は、変数ｖに、変数ｓｔｒのｉ番目の文字を代入する（ステップＳ１４０２）。

【０１４１】

共通辞書学習部２１１は、新規辞書（共通辞書７００）に、変数ｖの値が含まれるか調べる（ステップＳ１４０３）。

【０１４２】

ステップＳ１４０３で、変数ｖの値が含まれる場合（ＹＥＳ）、共通辞書学習部２１１は、変数ｄに１を加算、変数ｖに変数ｓｔｒのｉ番目からｉ＋ｄ番目の文字からなる文字列を代入し、符号ｋｅｙに変数ｖの値に対応する共通辞書７００または一時辞書７００’のｋｅｙ７０１’の値を代入し、ステップＳ１４０２に戻る（ステップＳ１４０４）。

【０１４３】

ステップＳ１４０３で、変数ｖの値が含まれない場合（ＮＯ）、共通辞書学習部２１１は、変数ｉにｉ＋ｄ＋１を代入し、変数ｄに０を代入し、新規辞書（共通辞書７００）に符号ｋｅｙ７０１がｋｅｙｍａｘ＋１で、Ｖａｌｕｅ７０２が変数ｖのエントリを追加し、統計カウンタ$[Key]^{++}$とする（ステップＳ１４０５）。

【０１４４】

共通辞書学習部２１１は、変数ｉの値と変数ｓｔｒの文字列の長さｌｅｎｇｔｈ（ｓｔｒ）と比較する（ステップＳ１４０６）。

【０１４５】

ステップＳ１４０６で、変数ｉが長さｌｅｎｇｔｈ（ｓｔｒ）よりも小さい場合（ＹＥＳ）、ステップＳ１４０３に戻る。

変数ｉが長さｌｅｎｇｔｈ（ｓｔｒ）よりも長い場合（ＮＯ）、共通辞書学習部２１１は、ヒット率の少ないｋｅｙを削除する（ステップＳ１４０７）。

【０１４６】

次に、共通辞書学習部２１１は、図７（Ａ）に示す新規辞書を共通辞書７００として共通辞書保存部２１９に保存する。

【０１４７】

以上述べた本実施例によれば、データを通信する端末に変更を施すことなく、通信データの内容が変化しても高効率に通信圧縮を行うことができる。また、通信コストの削減や通信速度の向上を実現することができる。

【実施例２】

【０１４８】

本実施例は、図１５に示すようにデータ転送装置１１０に学習データ選別部（学習データ選別プログラム）２２０を設け、デバイス情報が類似するデバイス（類似デバイス）を纏めて、学習に用いるデータの総量を削減するものである。

図１５～図１７を参照して、実施例１から変更がある部分のみ、以下に示す。

【０１４９】

図１５は、本発明のデータ転送装置をネットワークシステムに適用した他の構成例を示す図であって、データ転送装置１１０のソフトウェア（主記憶装置２０２）の論理的な接続関係を示す図、図１６は、学習データ選別部（学習データ選別プログラム）２２０の処理手順を示すフローチャート、図１７は、図１６のＳ１６０１１の詳細処理手順を示すフローチャートである。

【０１５０】

図１６のフローチャートに基づく動作は以下のとおりである。

学習データ選別部（学習データ選別プログラム）２２０は、類似デバイスが学習デバイスリストに登録済みか否かを判定する（ステップＳ１６０１）。

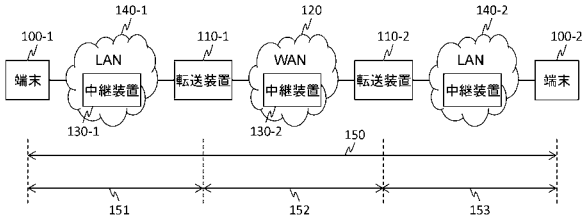
【０１５１】

学習データ選別部（学習データ選別プログラム）２２０は、類似デバイスが学習デバイスリストに登録済みの場合（ＹＥＳ）は、処理を終了し、登録済みでない場合（ＮＯ）は、類似デバイスを学習デバイスリストに登録する（ステップＳ１６０２）し、処理を終了

する。
【０１５２】
図１７のフローチャートに基づく動作は以下のとおりである。
図１８における通信先デバイスのデバイス情報１８０のＴｒａｆｉｃ１８０４、Ｃｏｍｐｒｅｓｓｉｏｎ１８０５、Ｈｉｔ　ｆｏｒ　ｎｔｈ　ｋｅｙ１８０６のそれぞれが予め設定した割合（Ｘ％）、つまり、類似デバイス割合以内のデバイスか否かを判定する（ステップＳ１６０１１～ステップＳ１６０１３）。
【０１５３】
すなわち、ステップＳ１６０１１にて、Ｔｒａｆｉｃ１８０４が予め設定したＸ％以内のデバイスがあり（ＹＥＳ）、ステップＳ１６０１２にて、そのデバイスのうち、Ｃｏｍｐｒｅｓｓｉｏｎ１８０５が予め設定したＸ％以内のデバイスがあり（ＹＥＳ）、かつ、ステップＳ１６０１３にて、Ｈｉｔ　ｆｏｒ　ｎｔｈ　ｋｅｙ１８０６が予め設定したＸ％以内のデバイスがある場合（ＹＥＳ）は、類似デバイス登録済みと判定し、上述した各ステップにて、それぞれ予め設定したＸ％以内のデバイスがない場合（ＮＯ）は、類似デバイスなしと判定する。
【０１５４】
例えば、図１８において、デバイスＩＤ「３」と「４」は、Ｔｒａｆｉｃ１８０４、Ｃｏｍｐｒｅｓｓｉｏｎ１８０５、Ｈｉｔ　ｆｏｒ　ｎｔｈ　ｋｅｙ１８０６のそれぞれについて予め設定した割合（Ｘ％）以内のデバイスであると判定した場合は、これらは類似デバイスと判定し、デバイスＩＤ「３」と「４」のデバイス情報は纏め、その一方のデバイスＩＤ［４］のデバイス情報を学習に用いるデータから削除する。
【０１５５】
本実施例によれば、例えば、類似デバイスによる類似通信が多いＩｏＴ等において、共通辞書を学習する際、類似デバイスの学習データを纏めることで学習量を低下させ、処理負荷を軽減することができる。
【符号の説明】
【０１５６】
１００　端末
１１０　データ転送装置
１２０　ＷＡＮ
１４０　ＬＡＮ
１３０　中継装置
２０１　演算装置
２０２　主記憶装置
２１１　共通辞書学習部
２１２　符号化部
２１３　復号化部
２１４　転送制御部
２１５　プロトコル処理部
２１６　フロー情報保存部
２１７　学習デバイスリスト保存部
２１８　学習用データ保存部
２１９　共通辞書保存部
２０３　二次記憶装置
２０４　ネットワークインターフェース
５００　フロー情報
６００　学習デバイスリスト
７００　共通辞書
７００’　一時辞書
１６０　デバイス情報

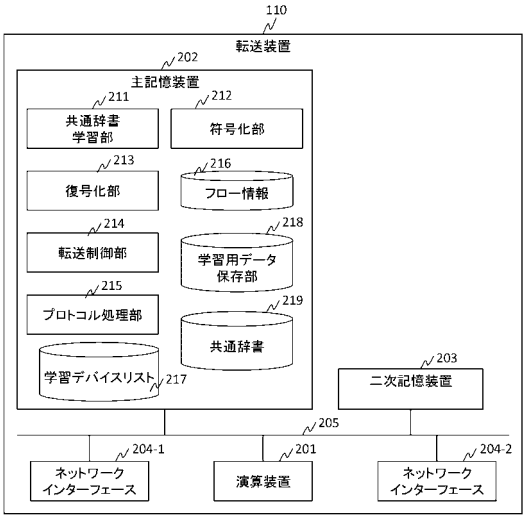
【図１】

図1



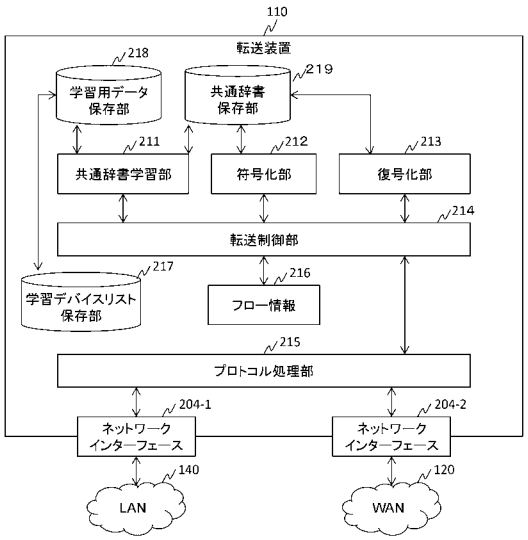
【図２】

図2



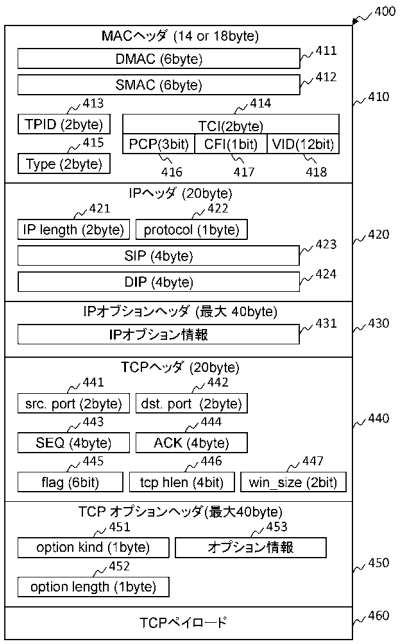
【図３】

図3



【図４】

図4



【図５】

図5

500

501 Flow-ID	502 VLAN	503 LAN-IP	504 WAN-IP	505 LAN-PORT	506 WAN-PORT
1	1	192.168.0.1	10.0.0.1	80	12345
2	10	192.169.0.2	123.123.123.123	8080	20842
3	10	192.170.0.3	111.111.111.111	80	45728

【図６】

図6

600

601 Device-ID	602 Device-IP
1	192.168.0.1
2	192.169.0.2
3	192.170.0.3

【図７（Ａ）】

図7A

700

701 Key	702 Value	703 統計カウンタ
0	<NULL>	0
1	a	14
2	b	223
3	aa	645
4	ba	3
5	aab	21

【図７（Ｂ）】

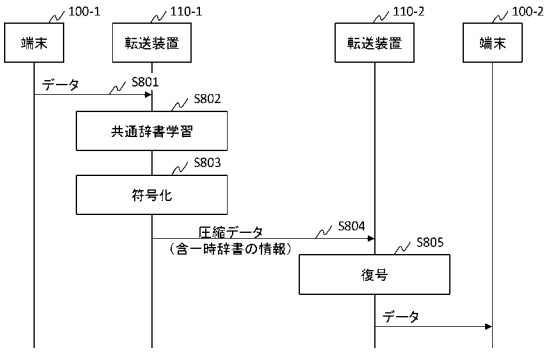
図7B

700'

701' Key	702' Value
6	aaa
7	ab

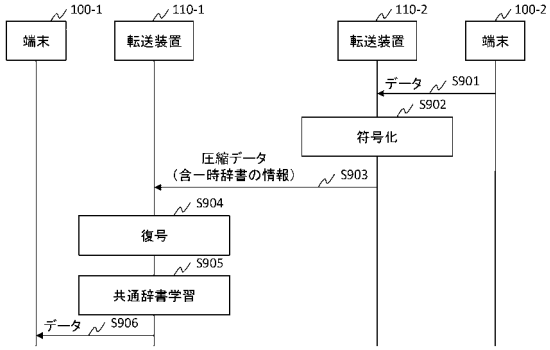
【図８】

図8



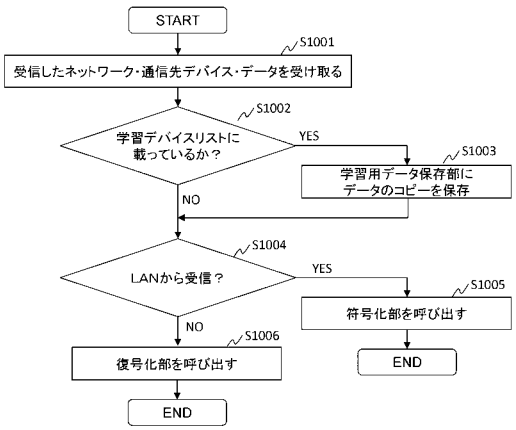
【図９】

図9



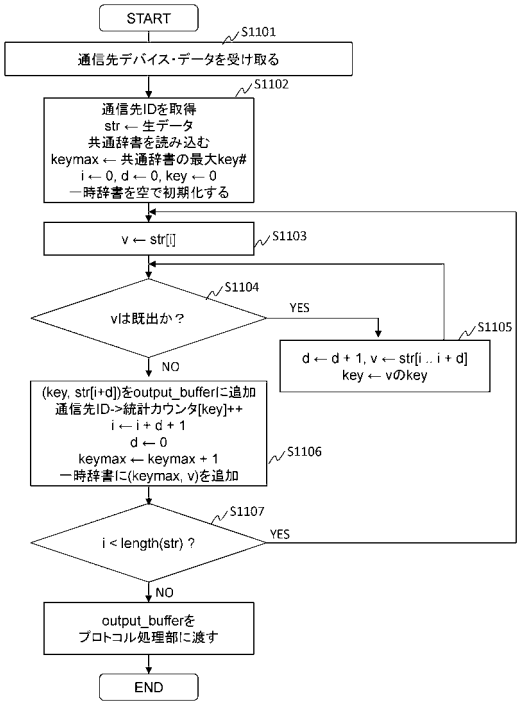
【図１０】

図10



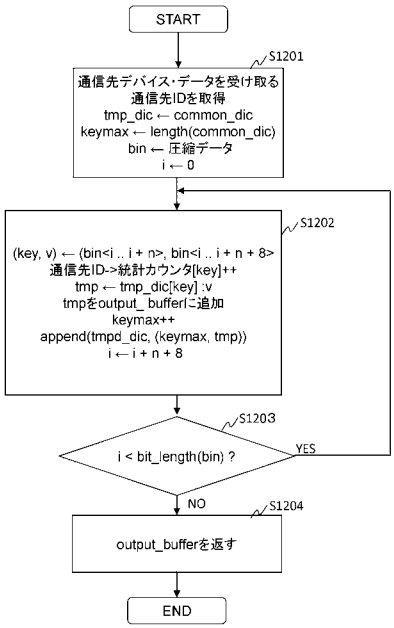
【図１１】

図11



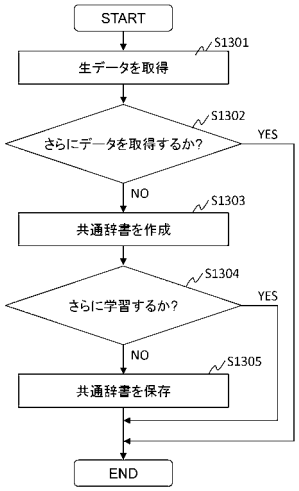
【図１２】

図12



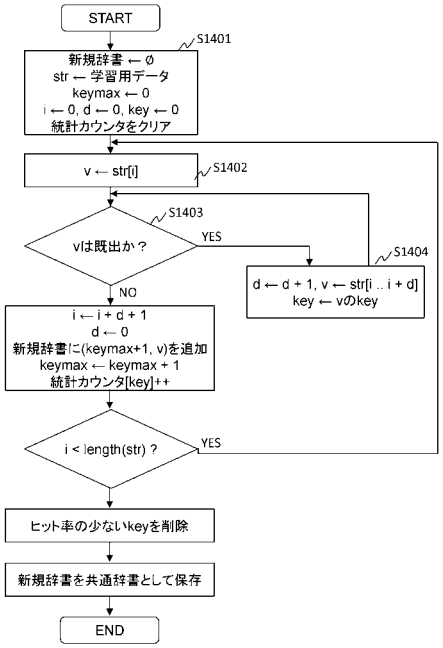
【図１３】

図13



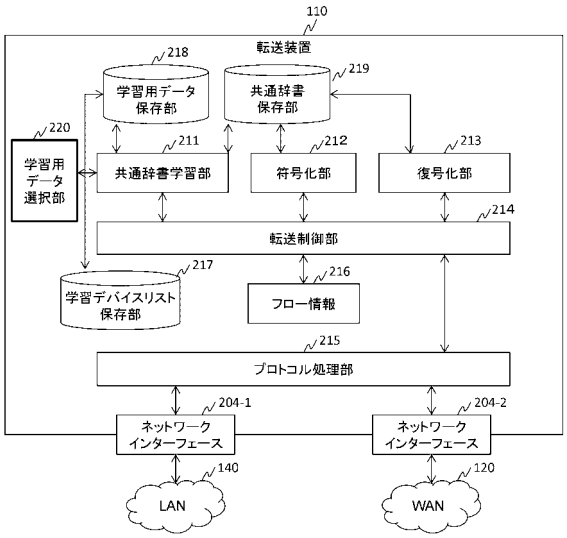
【図１４】

図14



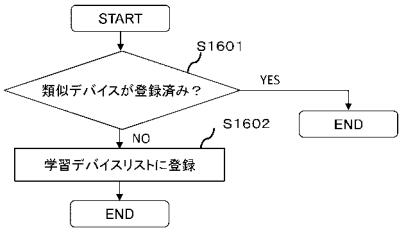
【図１５】

図15



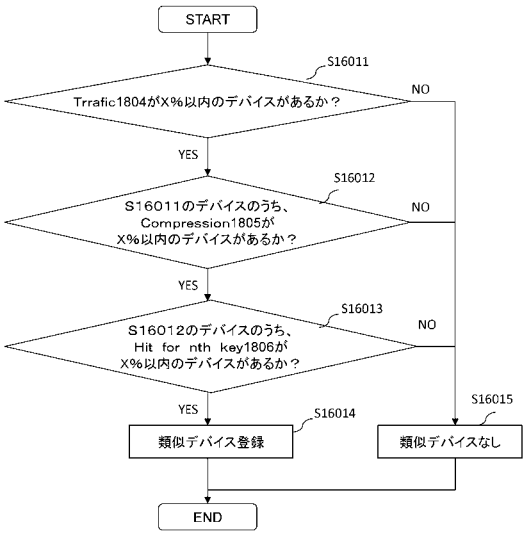
【図１６】

図16



【図１７】

図17



【図１８】

図18

180

1801 デバイスID	1802 VLAN	1803 IP	1804 Traffic	185 Compression	1806 Hit for nth key
1	1	192.168.0.1	100	70	900
2	10	192.169.0.2	30	50	20842
3	10	192.170.0.3	100	70	32

フロントページの続き

(72)発明者　仲　賢二
東京都千代田区丸の内一丁目６番６号　株式会社日立製作所内
(72)発明者　宮崎　悠樹
東京都千代田区丸の内一丁目６番６号　株式会社日立製作所内
(72)発明者　西本　絹之介
東京都千代田区丸の内一丁目６番６号　株式会社日立製作所内

Ｆターム(参考) 5J064 AA02 BC01 BD02
5K030 GA01 HA08 HD03 LA07 LB05