(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第4211282号
(P4211282)

(45)発行日　平成21年1月21日(2009.1.21)　(24)登録日　平成20年11月7日(2008.11.7)

(51)Int.Cl.		F I		
G06F 13/00	(2006.01)	G06F	13/00	520B
G06F 12/00	(2006.01)	G06F	13/00	520C
		G06F	13/00	520R
		G06F	12/00	545M

請求項の数 22　(全 17 頁)

(21)出願番号　特願2002-139232(P2002-139232)
(22)出願日　平成14年5月14日(2002.5.14)
(65)公開番号　特開2003-330619(P2003-330619A)
(43)公開日　平成15年11月21日(2003.11.21)
審査請求日　平成17年4月13日(2005.4.13)

前置審査

(73)特許権者　000002185
ソニー株式会社
東京都港区港南１丁目７番１号
(74)代理人　100067736
弁理士　小池　晃
(74)代理人　100096677
弁理士　伊賀　誠司
(74)代理人　100106781
弁理士　藤井　稔也
(74)代理人　100113424
弁理士　野口　信博
(74)代理人　100150898
弁理士　祐成　篤哉

最終頁に続く

(54)【発明の名称】データ蓄積方法及びデータ蓄積システム、並びに、データ記録制御装置、データ記録指令装置、データ受信装置及び情報処理端末

(57)【特許請求の範囲】

【請求項１】

ネットワークを介して接続された複数（ｎ個）の情報処理端末にデータを分散させて記録するデータ蓄積方法において、

上記複数の情報処理端末にデータを分散して記録する記録指令装置が、入力された第１のデータをｐ個のブロックに分割し符号化率ｑ／ｐでＦＥＣ（Forward Error Collection）符号化してｑ個の符号化ブロックを有する第２のデータに変換する符号化工程と、

上記記録指令装置が、上記第２のデータにおけるｑ個の符号化ブロックの各々に制御情報を付加してパケット化するパケット化工程と、

上記記録指令装置が、上記パケット化工程において生成されたパケットを上記複数の情報処理端末で記録する確率を表した記録確率αを生成する記録確率生成工程と、

上記記録指令装置が、上記パケットを上記複数の情報処理端末に送信する第１の送信工程と、

上記情報処理端末が、上記記録確率αで該パケットを記録する記録工程とを有し、

上記記録確率αで上記複数の情報処理端末の何れかに記録された上記パケットを送信指令装置が読み出す処理として、

上記送信指令装置が、上記情報処理端末が返信確率βで上記パケットを返信することを要求する返信要求パケットを生成する返信要求パケット生成工程と、

上記送信指令装置が、上記返信要求パケットを上記複数の情報処理端末に送信する第２の送信工程と、

上記複数の情報処理端末が、上記返信確率βで該パケットを返信する返信工程とを有し、

上記符号化率ｑ／ｐ、記録確率α、返信確率β、および上記情報処理端末の個数ｎを、ｐ≧ｑ×α×ｎ×βを満たすように設定する

データ蓄積方法。

【請求項２】

上記記録確率αは、上記パケットの上記制御情報に含まれ、該パケットにより上記複数の情報処理端末に送信され、

上記返信確率βは、上記返信要求パケットの制御情報に含まれ、該返信要求パケットにより上記複数の情報処理端末に送信される請求項１記載のデータ蓄積方法。

【請求項３】

上記記録確率αは、上記パケットとは別のパケットで送信され、

上記返信確率βは、上記返信要求パケットとは別のパケットで送信される請求項１記載のデータ蓄積方法。

【請求項４】

上記返信工程において返信されるパケット数が上記第１のデータのブロック数ｐと同じかそれ以上である請求項１記載のデータ蓄積方法。

【請求項５】

上記符号化率ｑ／ｐ、上記記録確率α、上記返信確率βを変更し上記パケットの分散を制御するパケット分散制御工程を有する請求項４記載のデータ蓄積方法。

【請求項６】

上記パケット分散制御工程では、上記記録確率αと、上記返信確率βと、上記情報処理端末の個数ｎと、第２のデータのブロック数ｑとの積が上記第１のデータのブロック数ｐと同じかそれ以上になるように各値を変更する請求項４記載のデータ蓄積方法。

【請求項７】

上記パケット分散制御工程では、上記ネットワークにおけるデータ損失確率ａと、上記記録確率αと、上記返信確率βと、上記情報処理端末の個数ｎと、第２のデータのブロック数ｑとの積が上記第１のデータのブロック数ｐと同じかそれ以上になるように各値を変更する請求項６記載のデータ蓄積方法。

【請求項８】

上記パケット分散制御工程では、上記複数の情報端末が応答しない確率ｂと、上記記録確率αと、上記返信確率βと、上記情報処理端末の個数ｎと、第２のデータのブロック数ｑとの積が上記第１のデータのブロック数ｐと同じかそれ以上になるように各値を変更する請求項４記載のデータ蓄積方法。

【請求項９】

上記記録確率生成工程および上記返信要求パケット生成工程は、

上記記録確率α、上記返信確率β、または上記符号化率ｑ／ｐを大きくすることで、上記返信工程において返信されるパケットが重複する確率を低くする請求項１記載のデータ蓄積方法。

【請求項１０】

ネットワークを介して接続された複数（ｎ個）の情報処理端末にデータを分散させて記録するデータ蓄積システムにおいて、

入力された第１のデータをｐ個のブロックに分割し符号化率ｑ／ｐでＦＥＣ（Forward Error Collection）符号化してｑ個の符号化ブロックを有する第２のデータに変換する符号化手段と、

上記第２のデータにおけるｑ個の符号化ブロックの各々に制御情報を付加してパケット化するパケット化手段と、

上記パケット化手段によって生成されたパケットを上記複数の情報処理端末で記録する確率を表した記録確率αを生成する記録確率生成手段と、

上記パケットを上記複数の情報処理端末に対して送出する送信手段と

を有するデータ記録指令装置と、
上記データ記録指令装置との間で上記パケットを送受する送受信手段と、
上記パケットを記録する記録手段と、
上記受信したパケットを上記記録確率αで上記記録手段に記録し、記録したパケットを返信確率βで返信するように制御する制御手段と
を有する複数の情報処理端末と、
上記情報処理端末との間でパケットを送受する送受信手段と、
上記情報処理端末の上記記録手段に記録されたパケットの返信を要求する返信要求パケットを生成する返信要求パケット生成手段と
を有する送信指令装置と
を備え、
これらのデータ記録指令装置、情報処理端末、及び送信指令装置がネットワークにより互いに接続されているデータ蓄積システム。

【請求項１１】
上記記録確率αは、上記パケットの上記制御情報に含まれ、該パケットにより上記複数の情報処理端末に送信され、
上記返信確率βは、上記返信要求パケットの制御情報に含まれ、該返信要求パケットにより上記複数の情報処理端末に送信される請求項１０記載のデータ蓄積システム。

【請求項１２】
上記記録確率αは、上記パケットとは別のパケットで送信され、
上記返信確率βは、上記返信要求パケットとは別のパケットで送信される請求項１０記載のデータ蓄積システム。

【請求項１３】
上記データ記録制御装置における受信手段にて受信されるパケット数が上記第１のデータのブロック数ｐと同じかそれ以上である請求項１０記載のデータ蓄積システム。

【請求項１４】
上記符号化率ｑ／ｐ、上記記録確率α、上記返信確率βを変更し上記パケットの分散を制御するパケット分散制御手段を有する請求項１３記載のデータ蓄積システム。

【請求項１５】
上記パケット分散制御手段は、上記記録確率αと、上記返信確率βと、上記情報処理端末の個数ｎと、第２のデータのブロック数ｑとの積が上記第１のデータのブロック数ｐと同じかそれ以上になるように各値を変更する請求項１３記載のデータ蓄積システム。

【請求項１６】
上記パケット分散制御手段は、上記ネットワークにおけるデータ損失確率ａと、上記記録確率αと、上記返信確率βと、上記情報処理端末の個数ｎと、第２のデータのブロック数ｑとの積が上記第１のデータのブロック数ｐと同じかそれ以上になるように各値を変更する請求項１５記載のデータ蓄積システム。

【請求項１７】
上記パケット分散制御手段は、上記複数の情報端末が応答しない確率ｂと、上記記録確率αと、上記返信確率βと、上記情報処理端末の個数ｎと、第２のデータのブロック数ｑとの積が上記第１のデータのブロック数ｐと同じかそれ以上になるように各値を変更する請求項１３記載のデータ蓄積システム。

【請求項１８】
上記記録確率生成手段および上記返信要求パケット生成手段は、
上記記録確率α、上記返信確率β、または上記符号化率ｑ／ｐを大きくすることで、上記複数の情報処理端末が返信するパケットが重複する確率を低くする請求項１０記載のデータ蓄積システム。

【請求項１９】
ネットワークを介して接続された複数（ｎ個）の情報処理端末にデータを分散して記録するデータ記録制御装置において、

入力された第１のデータをｐ個のブロックに分割し符号化率ｑ／ｐでＦＥＣ（Forward Error Collection）符号化してｑ個の符号化ブロックを有する第２のデータに変換する符号化手段と、
上記第２のデータにおけるｑ個の符号化ブロックの各々に制御情報を付加してパケット化するパケット化手段と、
上記パケット化手段によって生成されたパケットを上記複数の情報処理端末で記録する確率を表した記録確率αを生成する記録確率生成手段と、
上記情報処理端末に記録されたパケットを上記情報処理端末が返信する確率を示す返信確率βを含む制御情報が付加された返信要求パケットを生成する返信要求パケット生成手段と、
上記パケット及び上記返信要求パケットを上記複数の情報処理端末に対して送出する送信手段と、
上記複数の情報処理端末から送られるパケットを受信する受信手段と
を備えるデータ記録制御装置。
【請求項２０】
ネットワークを介して接続された複数（ｎ個）の情報処理端末にデータを分散して記録するデータ記録指令装置において、
入力された第１のデータをｐ個のブロックに分割し符号化率ｑ／ｐでＦＥＣ（Forward Error Collection）符号化してｑ個の符号化ブロックを有する第２のデータに変換する符号化手段と、
上記第２のデータにおけるｑ個の符号化ブロックの各々に制御情報を付加してパケット化するパケット化手段と、
上記パケット化手段によって生成されたパケットを上記複数の情報処理端末で記録する確率を表した記録確率αを生成する記録確率生成手段と、
上記パケットを上記複数の情報処理端末に対して送出する送信手段と
を備えるデータ記録指令装置。
【請求項２１】
ネットワークを介して接続された複数（ｎ個）の情報処理端末からデータを受信するデータ受信装置において、
上記情報処理端末に記録されたパケットを上記情報処理端末が返信する確率を示す返信確率βを含む制御情報が付加された返信要求パケットを生成する返信要求パケット生成手段と、
上記返信要求パケットを上記複数の情報処理端末に対して送出する送信手段と、
上記複数の情報処理端末から送られるパケットを受信する受信手段と
を備えるデータ受信装置。
【請求項２２】
ネットワークで接続された外部装置との間でパケット化されたデータを送受する送受信手段と、
上記パケットを記録する記録手段と、
上記パケットを該パケットの付加情報に記述された記録確率αで記録するか否か、および記録したパケットを返信確率βで返信するか否かを制御する制御手段と
を備える情報処理端末。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】
本発明は、データ蓄積方法及びデータ蓄積システム、並びにデータ記録制御装置、データ受信装置及び情報処理端末に関し、ネットワークを介して接続されたノード間でデータを分散して蓄積するデータ蓄積方法及びデータ蓄積システムに関する。また、このデータ蓄積システムに用いて好適なデータ記録制御装置、データ記録指令装置、データ受信装置及び情報処理端末に関する。

【０００２】
【従来の技術】
近年、あるデータを、ネットワークを介して互いに接続された多数の情報処理端末に分散して記録する大規模ストレージシステムが注目されている。このような分散型ストレージシステムにおいて、データを記録管理するサーバは、マルチキャスト等によって情報処理端末やその他のサーバにデータを送信して、データを情報処理端末や他のサーバに備えられたローカルの記録媒体に記録している。
【０００３】
この場合、オンデマンドでデータを取り出せるようにするためには、記録媒体に多量のデータを記録しなければならない。例えば、１本当たり約２ギガバイトのデータ容量になる映画の場合、このような映像データを５００本分記録するとすれば、１テラバイト以上の容量が必要となる。
【０００４】
また、ストリーミングによってデータを提供する場合の例として、サーバがデータを要求しているクライアントに対してユニキャストでデータを提供する際には、エラーのない伝送を行うために、例えば、ＴＣＰ／ＩＰの到着済み信号（ＡＣＫ）のようにデータの再送を要求するプロトコルが用いられる。
【０００５】
ところが、この手法は、サーバ側に多大な負担がかかるため、高性能なサーバ１台を用いたとしても、現状では、数百台のクライアントにしかサービスを提供することができない。また、ＵＤＰ／ＩＰのようなＡＣＫを用いないプロトコルを使用したとしても、サービス可能なクライアントの数は、数千台程度である。このように、ストリーミングによってデータを提供しようとすると、サーバ側のコストが増大し、クライアントの数が制限されてしまう。
【０００６】
そこで近年では、マルチキャスト技術にＦＥＣ（Forward Error Correction）を用いて、データの再送を要求することなく複数のクライアントにデータを送信する方式が提案されている。これは、サーバがマルチキャストでストリームを繰り返し送信し、クライアントは、このストリームから必要な信号を拾い上げ、拾い上げたデータを復号して再生する方式である。
【０００７】
【発明が解決しようとする課題】
この方式を利用して、１本２ギガバイトになる映画の映像データ５００本分を１０分以内に送信する場合には、約１４．７ギガビット／秒の伝送帯域が必要である。さらに、同量の映像データを１分以内に送信する場合には、約１４７ギガビット／秒の伝送帯域が必要になる。これは、理論値であるが、このような容量及び伝送方式に耐えうるサーバは、非常にコストがかかり、実現したとしても実用的でない。また、複数のホストにデータを分散して記録するという方式もあるが、このシステムを実現しようとすると、巨大なデータを複数のサーバで管理しなければならないため、データ管理やデータ通信のための処理が増大してしまう。
【０００８】
そこで本発明は、上述したような従来の実情に鑑みて提案されたものであり、複雑な処理や膨大な伝送帯域を必要とせず、個々の端末に負担をかけることなく多量のデータを分散して管理できるデータ蓄積方法及びデータ蓄積システム、並びに、複雑な処理や膨大な伝送帯域を必要とせず、個々の端末に負担をかけることなく多量のデータを分散して記録するとともに記録場所からデータを取り出すよう制御するデータ記録制御装置、複雑な処理や膨大な伝送帯域を必要とせず、個々の端末に負担をかけることなく多量のデータを分散して記録するデータ記録指令装置、個々の端末に分散して記録されたデータを受信するデータ受信装置及び多量のデータを分散して記録できる情報処理端末を提供することを目的とする。

【０００９】
【課題を解決するための手段】
上述した目的を達成するために、本発明に係るデータ蓄積方法は、ネットワークを介して接続された複数（ｎ個）の情報処理端末にデータを分散させて記録するデータ蓄積方法において、上記複数の情報処理端末にデータを分散して記録する記録指令装置が、入力された第１のデータをｐ個のブロックに分割し符号化率ｑ／ｐでＦＥＣ（Forward Error Collection）符号化してｑ個の符号化ブロックを有する第２のデータに変換する符号化工程と、上記記録指令装置が、上記第２のデータにおけるｑ個の符号化ブロックの各々に制御情報を付加してパケット化するパケット化工程と、上記記録指令装置が、上記パケット化工程において生成されたパケットを上記複数の情報処理端末で記録する確率を表した記録確率αを生成する記録確率生成工程と、上記記録指令装置が、上記パケットを上記複数の情報処理端末に送信する第１の送信工程と、上記情報処理端末が、上記記録確率αで該パケットを記録する記録工程とを有し、上記記録確率αで上記複数の情報処理端末の何れかに記録された上記パケットを送信指令装置が読み出す処理として、上記送信指令装置が、上記情報処理端末が返信確率βで上記パケットを返信することを要求する返信要求パケットを生成する返信要求パケット生成工程と、上記送信指令装置が、上記返信要求パケットを上記複数の情報処理端末に送信する第２の送信工程と、上記複数の情報処理端末が、上記返信確率βで該パケットを返信する返信工程とを有し、上記符号化率ｑ／ｐ、記録確率α、返信確率β、および上記情報処理端末の個数ｎを、ｐ≧ｑ×α×ｎ×βを満たすように設定する。
【００１０】
また、本発明に係るデータ蓄積システムは、ネットワークを介して接続された複数（ｎ個）の情報処理端末にデータを分散させて記録するデータ蓄積システムにおいて、入力された第１のデータをｐ個のブロックに分割し符号化率ｑ／ｐでＦＥＣ（Forward Error Collection）符号化してｑ個の符号化ブロックを有する第２のデータに変換する符号化手段と、上記第２のデータにおけるｑ個の符号化ブロックの各々に制御情報を付加してパケット化するパケット化手段と、上記パケット化手段によって生成されたパケットを上記複数の情報処理端末で記録する確率を表した記録確率αを生成する記録確率生成手段と、上記パケットを上記複数の情報処理端末に対して送出する送信手段とを有するデータ記録指令装置と、上記データ記録指令装置との間で上記パケットを送受する送受信手段と、上記パケットを記録する記録手段と、上記受信したパケットを上記記録確率αで上記記録手段に記録し、記録したパケットを返信確率βで返信するように制御する制御手段とを有する複数の情報処理端末と、上記情報処理端末との間でパケットを送受する送受信手段と、上記情報処理端末の上記記録手段に記録されたパケットの返信を要求する返信要求パケットを生成する返信要求パケット生成手段とを有する送信指令装置とを備え、これらのデータ記録指令装置、情報処理端末、及び送信指令装置がネットワークにより互いに接続されている。
【００１１】
また、上述した目的を達成するために、本発明に係るデータ記録制御装置は、ネットワークを介して接続された複数（ｎ個）の情報処理端末にデータを分散して記録するデータ記録制御装置において、入力された第１のデータをｐ個のブロックに分割し符号化率ｑ／ｐでＦＥＣ（Forward Error Collection）符号化してｑ個の符号化ブロックを有する第２のデータに変換する符号化手段と、上記第２のデータにおけるｑ個の符号化ブロックの各々に制御情報を付加してパケット化するパケット化手段と、上記パケット化手段によって生成されたパケットを上記複数の情報処理端末で記録する確率を表した記録確率αを生成する記録確率生成手段と、上記情報処理端末に記録されたパケットを上記情報処理端末が返信する確率を示す返信確率βを含む制御情報が付加された返信要求パケットを生成する返信要求パケット生成手段と、上記パケット及び上記返信要求パケットを上記複数の情報処理端末に対して送出する送信手段と、上記複数の情報処理端末から送られるパケットを受信する受信手段とを備える。

【００１２】
また、上述した目的を達成するために、本発明に係るデータ記録指令装置は、ネットワークを介して接続された複数（ｎ個）の情報処理端末にデータを分散して記録するデータ記録指令装置において、入力された第１のデータをｐ個のブロックに分割し符号化率ｑ／ｐでＦＥＣ（Forward Error Collection）符号化してｑ個の符号化ブロックを有する第２のデータに変換する符号化手段と、上記第２のデータにおけるｑ個の符号化ブロックの各々に制御情報を付加してパケット化するパケット化手段と、上記パケット化手段によって生成されたパケットを上記複数の情報処理端末で記録する確率を表した記録確率αを生成する記録確率生成手段と、上記パケットを上記複数の情報処理端末に対して送出する送信手段とを備える。
【００１３】
また、本発明に係るデータ受信装置は、ネットワークを介して接続された複数（ｎ個）の情報処理端末からデータを受信するデータ受信装置において、複数の情報処理端末に記録されたパケットを返信する確率を示す返信確率βを含む制御情報が付加された返信要求パケットを生成する返信要求パケット生成手段と、返信要求パケットを複数の情報処理端末に対して送出する送信手段と、複数の情報処理端末から送られるパケットを受信する受信手段とを備える。
【００１４】
また、本発明に係る情報処理端末は、ネットワークで接続された外部装置との間でパケット化されたデータを送受する送受信手段と、パケットを記録する記録手段と、パケットを該パケットの付加情報に記述された記録確率αで記録するか否か、および記録したパケットを返信確率βで返信するか否かを制御する制御手段とを備える。
【００１５】
【発明の実施の形態】
以下、本発明を適応した具体例について図を参照して詳細に説明する。図１は、ネットワークを構成する端末にデータを分散して記録する分散型ストレージシステムの具体例を示す図である。分散型ストレージシステムは、網の目のように接続されたｎ個のノード20_1、20_2、…、20_{n-1}、20_nと、各ノードへのデータの記録を指示制御する記録指令装置１０と、各ノードに記録されたデータを読み出す送信指令装置３０とを有する。本具体例では、記録指令装置１０と、ノード２０と、送信指令装置３０とを別の装置として説明するが、これら両装置の機能を有する記録送信指令装置のような装置であってもよい。また、ノードの各々に記録指令装置１０や送信指令装置３０としての機能を装備することもできる。この場合、ネットワークを構成する各装置を区別な使用できる。
【００１６】
また、図１では、ネットワークを構成するルータなどの伝送制御装置を省略してあるが、実際には、ノード２０を通過するパケットの経路を選択するルータなどの伝送制御装置が設けられている。伝送制御装置は、ノード２０は別に設けられていてもよいし、ノード２０が伝送制御装置のような機能を有していてもよい。
【００１７】
次に、記録指令装置１０について説明する。図２は、記録指令装置１０の内部構成を示す図である。記録指令装置１０は、ＦＥＣ（Forward Error Collection）符号化を施すＦＥＣエンコーダ１１、符号化したデータをインターリーブするインターリーバ１２、インターリーバ１２から出力されたデータをパケットに変換するパケット生成部１３、及びネットワークとの接続を行うネットワークインターフェース１４を有しており、データ入力部１５を介してデータが入力される。
【００１８】
ＦＥＣエンコーダ１１は、ＦＥＣ符号化を用いて入力されたデータを符号化する。ここで、ＦＥＣ符号化とは、トルネード符号化方式、リードトルネード符号化方式、ターボ符号方式などの受信側で誤り訂正を行う符号化方式の総称であり、ＦＥＣエンコーダ１１は、データ入力部１５から入力されたデータをｐ個のブロックに分割し、このｐ個のブロック

にＦＥＣ符号化を施してｑ個のブロックに変換している。このｐ個のブロックからｑ個のブロックに符号化することを符号化率ｑ／ｐの符号化といい、この符号化率ｑ／ｐを変更することによって、この分散型ストレージシステムの記録効率や伝送効率を変更することができる。

【００１９】

インターリーバ１２は、符号化されたデータの順番を並び換える。インターリーバ１２は、このようにインターリーブすることによってデータを分散させ、パケットの消失によって発生するバーストエラーがランダムエラーになるようにすり替える。

【００２０】

パケット生成部１３は、インターリーバ１２から送信されたデータを所定のサイズに分割し、この分割したデータに、制御情報を付加してパケット化する。図３は、パケット生成部１３によって生成されるパケット４０の構造を示す図である。パケット４０は、ヘッダ４１、記録確率記述部４２、ペイロード４３、フッタ４４から構成される。ペイロード４３には、ＦＥＣエンコーダ１１によって変換された送信情報が記述されている。ヘッダ４１とフッタ４４には、データの種類を示すデータＩＤ、ＣＲＣ（Cyclic Redundancy Check）のチェックサム、ノード２０のＧＵＩＤ（Global Unique ID）、ネットワークアドレスなどの制御情報が記述されている。

【００２１】

記録確率記述部４２には、後述する各ノードがこのパケットを記録する確率が記述されている。ノードの制御部２２は、この記録確率αに基づいてパケットを記録する。

【００２２】

分散型ストレージシステムを構成する全てのノードは、この記録確率に基づいてパケットを記録するか否かを決定する。これにより、分散型ストレージシステムを構成するノードに確率αでデータが記録されることになる。この分散型ストレージシステムでは、ノードの数ｎが十分に大きく、且つ復号されたブロック数ｑが十分に大きい場合に、各ノードに均等な確率でデータを分散することができる。

【００２３】

ネットワークインターフェース１４は、パケット化されたデータを入力し、ユニキャスト又はマルチキャストで各ノード２０にデータを送信する。

【００２４】

次に、ノード２０の構成について説明する。図４は、ノード２０の構成を示す図である。ノード２０は、ネットワークインターフェース２１、制御部２２、記録部２３を有し、ネットワークインターフェース２１を介してネットワークと接続している。

【００２５】

制御部２２は、入力したパケット４０を記録部２３に記録する記録処理と、他のノード２０からの要求に応じてデータを返信する返信処理を実行する。

【００２６】

記録処理は、入力したパケット４０の記録確率αをもとに、記録部２３にパケット４０を記録するか否かを判定する処理であり、上述したように、分散型ストレージシステムを構成するノードの数ｎが十分に大きく、且つ復号化されたブロックの個数ｑが十分に大きい場合、全てのノードにパケットが均等に記録され、分散型ストレージシステム全体としてαの確率でデータが記録される。

【００２７】

返信処理は、送信指令装置から送られたデータの送信要求をするパケットを受信すると開始する処理である。このデータ送信要求パケットのデータ構造については後述するが、このパケットには、データを返信する確率βが記述されており、制御部２２は、記録されたデータを返信確率βで返信する。ここで、データを返信すると判定したとき、制御部２２は、目的のデータを検索して返信パケットを生成する。

【００２８】

次に、送信指令装置３０について説明する。図５は、送信指令装置３０の構成を示す図で

ある。送信指令装置３０は、ネットワークを介しての外部とのデータの送受信を行うネットワークインターフェース３１、ノードにデータの送信を要求するデータ要求部３２、パケットに分割されたデータを結合するパケット結合部３３、デインターリーバ３４、ＦＥＣデコーダ３５を有し、復号したデータは応用処理部３６に入力され、モニタやスピーカ（図示省略）などの外部機器に出力したり、図示しない記録装置に格納されたりする。

【００２９】

データ要求部３２は、分散型ストレージシステムを構成する各ノードにデータを要求するパケットを送信する。図６は、データを要求するパケット５０の構成を示す図である。パケット５０は、ヘッダ５１、返信確率記述部５２、リクエスト記述部５３、フッタ５４から構成される。リクエスト記述部５３には、要求するデータを識別するためのデータＩＤが記録される。ヘッダ５１とフッタ５４には、ＣＲＣのチェックサム、ノードのネットワークアドレスやＧＵＩＤ、データの順序を示すシーケンス番号などの制御情報が記録される。

【００３０】

返信確率記述部５２には、返信確率βが記述されている。返信確率βは、パケット５０を受信したノードがデータを返信するか否かの判定を行うための変数である。この変数をもとに、データを返信すると判定するノードもあれば、データを返信しないと判定するノードもあるが、返信確率βは、分散型ストレージシステム全体をマクロ的にみたときの値であり、分散型ストレージシステム全体では、各ノードのデータを返信する確率がβになる。そのため、分散型ストレージシステムにｎ個のノード２０が存在する場合に、返信されるパケットの割合は、ノード２０の個数ｎと返信確率βを掛け合わせた値ｎ×βとなる。

【００３１】

パケット結合部３３は、各ノード２０から返信されたパケットを結合する。図７は、ノードから返信されたパケット６０のデータ構造を示す図である。図７に示すようにパケット６０は、ヘッダ６１、ペイロード６２、フッタ６３から構成され、ペイロード６２には送信するデータが格納され、ヘッダ６１とフッダ６３には、ＣＲＣのチェックサムや受信側のノードのネットワークアドレス、パケットの順序を示すシーケンス番号など、制御情報が格納される。

【００３２】

図７のパケット６０を受信すると、パケット結合部３３は、シーケンス番号を読み取り、受信したパケット６０の順序を入れ替え、ヘッダ６１やフッタ６３などの制御情報を除去して、シーケンス番号の順にパケットを結合する。

【００３３】

デインターリーバ３４は、結合されたパケットにデインターリーブをかけ、データの並びを整列させる。ＦＥＣデコーダ３５は、デインターリーブされたデータにＦＥＣ復号を施し、元のデータを復元する。

【００３４】

ＦＥＣデコーダ３５によって復号されたデータは、応用処理部３６に出力される。応用処理部３６は、復号されたデータを、図示しない記録部に保存したり、モニタやスピーカなどの入出力インターフェースに出力する。

【００３５】

このように、本具体例における分散型ストレージシステムは、記録確率αで各ノードにデータを記録し、各ノードに記録したデータを返信確率βで返信させるシステムであり、記録指令装置１０から出力される元データは、α×ｎ×βの割合で返信される。例えば、ｐ個のブロックをｑ個のブロックに符号化すると、ｑ×α×ｎ×β個のブロックが返信される。後述する論文ＲＩＺＺ０９７に記載のように、この返信されたブロックの個数が復号前のブロックの個数ｐよりも多い場合、データは復号可能である。そのため、返信されるブロックの個数がｐ個より多くなるように、α、β、ｑ／ｐの値を決定しておくと、目的のデータを復号することができる。

【００３６】

次に、分散型ストレージシステムの動作について、図８～図１２にしたがって説明する。まず、データの記録動作について説明する。データを記録する際、記録指令装置１０は、データ入力部１５から目的のデータを入力し、入力したデータをＦＥＣエンコーダ１１に出力する。

【００３７】

ＦＥＣエンコーダ１１は、図８（ａ）に示すように、入力したデータをｐ個のブロックに分割する。そして、符号化率ｑ／ｐのＦＥＣ符号化を施し、ｐ個のブロックに分割したデータをｑ個の符号化ブロックに変換する。

【００３８】

上述したように、ＦＥＣ符号化とは、トルネード符号化方式、リードトルネード符号化方式、ターボ符号化方式などの受信側で誤り訂正を行う符号化方式の総称であり、ＦＥＣ符号化を用いて、あるデータを符号化率ｑ／ｐで符号化した場合、論文ＲＩＺＺ０９７（http://www.iet.unipi.it/~luigi/fec.html#fec.ps)に発表されているように、ｐ個以上の符号化されたブロックが残存すれば、幾つかのブロックが消失しても、元のメッセージが復元できるようになっている。

【００３９】

ＦＥＣ符号化を施されたデータは、インターリーバ１２に出力され、インターリーバ１２は、符号化されたデータの順番を並び換え、データを分散させる。インターリーバ１２を通過したデータは、パケット生成部１３に出力される。パケット生成部１３は、分散されたデータを所定の大きさに分割し、分割したデータにヘッダやフッタを付加して、図８（ｂ）に示すようなパケット４０を生成する。

【００４０】

このように、生成されたパケット４０は、ネットワークインターフェース１４に出力される。ネットワークインターフェース１４は、図９に示すように、ユニキャスト若しくはマルチキャストを用いて、パケット化したデータを、分散型ストレージシステムを構成する各ノードに送信する。

【００４１】

記録指令装置１０から送信されたパケット４０を受信したノードは、記録処理を開始し、パケット４０の記録確率αを読み出して、パケット４０を記録する否かの判定を行う。そして、記録すると判定した場合には、データを記録部２２にデータを記録し、記録しないと判定した場合には、受信したパケット４０をその他のノードに出力する。記録確率αは、分散型ストレージシステム全体で、どの程度のデータを記録するかを示す値であり、送信データがｑ個の符号化ブロックであるとき、分散型ストレージシステム全体には、ｑ×α個のデータが記録される。

【００４２】

次に、データを取り出す処理について説明する。データを取り出すとき、送信指令装置３０は、パケット５０を生成し、図１０に示すように、ユニキャスト若しくはマルチキャストを用いて、各ノードにパケット５０を出力する。

【００４３】

パケット５０を受信したノードは、返信処理を開始する。パケット５０には、通常の制御情報の他に、データを識別するデータＩＤ、パケットを返信する返信確率βなどが記録されており、返信処理では、パケット５０の返信確率βをもとにデータを返信するか否かの判定する。ここで、データを返信すると判定した場合、ノードの制御部２２は、データを包含したパケット６０を生成し、送信指令装置３０に返信する。また、データを返信しないと判定した場合には、ノードの制御部２２は、何も実行しない。

【００４４】

図１１は、各ノードから送信指令装置３０にパケットが送信される様子を示した図である。図１１に示すように、分散型ストレージシステムには、データを返信するノードとそうでないノードがあるが、全体でみると、各ノードは返信確率βの割合でデータを返信するようになっており、分散型ストレージシステムにｎ個のノードが存在すると、ｎ×βの割

合のデータが返信されることになる。上述したように、分散型ストレージシステムには、記録確率αをもとにｑ×αのデータが記録されているので、ｑ×α×ｎ×β個のデータが返信される。

【００４５】

次に、図１２を参照して、データの復号化について説明する。各ノードから送信されるデータを受信した送信指令装置３０は、受信したデータをパケット結合部３２に入力し、パケットの結合を行う。図１２（ａ）は、ノードから送信指令装置３０に返信されるパケット６０を示す図である。図１２（ａ）において、斜線で描かれた箇所がデータの損失された箇所を示す。分散型ストレージシステムでは、元のデータをα×ｎ×βの確率で返信するため、図１２（ａ）に示すように、全てのパケットが返信される訳ではなく、幾つかのパケットの消失が起きている。

【００４６】

しかしながら、このデータはインターリーブされているため、デインターリーブをかけると、バースト的なデータの消失は、分散し、ランダムなデータの消失に変換される。

【００４７】

上述したように、ＦＥＣ符号化では、符号化率ｑ／ｐでデータを符号化した場合、符号化されたデータがｐブロック以上存在すれば、データを復号することができる。送信したｑ個のブロックがα×ｎ×βの割合で返信されたときは、返信されるデータの個数は、ｑ×α×ｎ×βである。この個数ｑ×α×ｎ×βがｐ個以上であるようにｑ／ｐ、α、βの値を設定すれば、元のデータを復元できる。

【００４８】

また、本具体例における分散型ストレージシステムは、ｐ≧ｑ×α×ｎ×βを満たすように、符号化率ｑ／ｐ、記録確率α、返信確率βを設定すればよいので、上述の式を満たす範囲で符号化率ｑ／ｐ、記録確率α、返信確率βを変更することにより、データの記録効率や伝送効率を変更させることができる。以下、ｑ／ｐ、α、βの各パラメータの設定例について説明する。

【００４９】

例えば、非常に多くの返信要求があるデータに対して、記録確率αの値を大きくし、返信確率βの値を小さくすると、各ノードから送信されるデータが少なくなり、ノードにおけるデータの検索処理やデータの送信処理が簡略化されるようになる。

【００５０】

また、記録確率αの値を大きくするかわりに、符号化率ｑ／ｐの値を大きくして、返信確率βを小さくしても、各ノードにおけるデータの検索処理やデータの送信処理を簡略化することができる。

【００５１】

また、符号化率ｑ／ｐを小さくし、記録確率αを大きくすると、送信するパケットの数を抑えることができる。これは、ｐが十分大きいときに効果的である。また、記録確率αを小さくし、符号化率ｑ／ｐを小さくすることで、同一のパケットを複数のノードに記録することを避けることができる。これは、ｐが十分小さいときに効果がある。

【００５２】

また、返信確率βを大きくし、記録確率α又は符号化率ｑ／ｐを小さくすることにより、分散型ストレージシステム全体に記録される符号化データの容量を小さくすることができる。或いは、データの記録時、出力時又は送信時などに、パケットの消失する確率をａとすると、ａ×ｎ×α×ｑ×βがｐよりも十分大きくなるようにα、β、ｑ／ｐの値を制御することによって、十分な数のデータが返信される。

【００５３】

また、複数のノードから返信されるユニークなパケットの個数を数学的に推定し、記録確率、返信確率、または符号化率を大きくすることで、ユニークなパケットが到着する確率を高くすることができる。

【００５４】

以上のように、本具体例における分散型ストレージシステムは、システムを構成するノードに記録確率αを記述したパケット４０を送信することにより、データを分散して記録する。そして、ノードに記録されたデータを返信確率βで返信させることにより、データの取り出しを行う。このように、データを分散して記録すると、一つのサーバにデータ管理の負荷が集中することなく、データを格納することができる。また、本具体例における分散型ストレージシステムでは、一つのデータを複数のノードで共有するため、システム全体で必要となるデータ容量が少なくて済む。

【００５５】

また、複数のノードからデータを受信するようにすると、一つのサーバにトラフィックが集中することがなくなり、安定した通信量でデータの送受信が行える。

【００５６】

また、α、β、ｑ／ｐの値を変更することにより、伝送効率や記録するデータ量を変更することができる。さらに、データの記録時、出力時、送信時などにパケットが消失するとき、消失する確率ａを考慮して、α、β、ｑ／ｐのパラメータの値を変更して、パケットが消失しても復号に十分なデータを取り出せるようにすることもできる。

【００５７】

また、この分散型ストレージシステムでは、記録確率と返信確率に基づいた演算によってデータの分散記録・読み出しを実行できるため、データの管理が単純であり、カムコーダや携帯電話などの処理能力の少ない家庭用の機器であってもこのシステムに適応できる。また、カムコーダや携帯電話などの処理能力の低い家庭用の機器をノードとして用いることができるようになるため、数百万台規模の分散型ストレージシステムが容易に構築できる。

【００５８】

また、上記分散ストレージシステムにおいては、記録指令装置は、記録させるデータと記録確率を同じパケットで送信したが、記録させるデータと記録確率を異なるパケットで送信したり、外部の記録装置に記録し、各情報処理ノードから参照するようにしてもよい。

【００５９】

なお、上述した具体例は、発明の一例に過ぎず、本発明の要旨を含む変形例は、本発明に含まれるものとする。例えば、ノードとしては、カムコーダ、パーソナルビデオレコーダーやホームゲートウェイなどが考えられるが、データを記録する記録部と、所定の演算を行う制御部と、データの送受信を行うネットワークインターフェースを有していればその他の構成を備える装置であってもよい。

【００６０】

また、記録確率α、返信確率βをパケット中に記録したが、記録確率α、返信確率βを任意の記録装置、若しくはパケットなどに記録し、各ノードがその値を参照するようにしてもよい。特に、ＦＥＣ符号化としてリードトルネード符号化方式を利用した場合には、インターリーブ処理を省略することもできる。

【００６１】

【発明の効果】

以上詳細に説明したように、本発明に係るデータ蓄積方法によれば、第１のデータを符号化して得た第２のデータをパケット化し、個々のパケットを複数の情報処理端末で記録するか否かを表す記録確率を生成し、記録確率とパケットを複数の情報処理端末に送り出し、この記録確率で複数の情報処理端末の何れかに記録されたパケットを読み出す際には、パケットを返信するか否かを示す返信確率を付して返信要求データを生成して返信要求データを複数の情報処理端末に送信し、複数の情報処理端末は、返信確率に基づいて該パケットを返信することにより、複雑な処理や膨大な伝送帯域を必要とせず、個々の端末に負担をかけることなく多量のデータを分散して記録するとともに記録場所からデータを取り出して再構成することができる。

【００６２】

また、本発明に係るデータ蓄積システムによれば、第１のデータを符号化して得た第２の

データをパケット化し、個々のパケットを複数の情報処理端末で記録するか否かを表す記録確率を生成し、パケットおよび記録確率を送り出し、この記録確率で複数の情報処理端末の何れかに記録されたパケットを読み出す際には、パケットを返信するか否かを示す返信確率を付して返信要求データを生成して返信要求データを複数の情報処理端末に送信し、複数の情報処理端末は、返信確率に基づいて該パケットを返信することにより、複雑な処理や膨大な伝送帯域を必要とせず、個々の端末に負担をかけることなく多量のデータを分散して記録するとともに記録場所からデータを取り出して再構成することができる。

【００６３】

また、本発明に係るデータ記録制御装置によれば、複雑な処理や膨大な伝送帯域を必要とせず、個々の端末に負担をかけることなく多量のデータを分散して記録するとともに記録場所からデータを取り出すよう制御し、これを再構成できる。

【００６４】

また、本発明に係るデータ記録指令装置によれば、複雑な処理や膨大な伝送帯域を必要とせず、個々の端末に負担をかけることなく多量のデータを分散して記録することができる。

【００６５】

さらに、本発明に係るデータ受信装置によれば、複雑な処理や膨大な伝送帯域を必要とせず、個々の端末に分散して記録されたデータを受信して再構成することができる。

【００６６】

さらにまた、本発明に係る情報処理端末によれば、複雑な処理や膨大な伝送帯域を必要とせず、負担なく多量のデータを分散して記録できる。

【図面の簡単な説明】

【図１】分散型ストレージシステムの構成を示す図である。

【図２】記録指令装置の内部構成を示す図である。

【図３】記録するデータを格納するパケットの構成を示す図である。

【図４】ノードの内部構成を示す図である。

【図５】送信指令装置の内部構成を示す図である。

【図６】データの送信を要求するパケットの構成を示す図である。

【図７】ノードから返信されるパケットの構成を示す図である。

【図８】パケットの生成工程を示す図である。

【図９】記録するデータを含むパケットの送信経路を示す図である。

【図１０】データの返信を要求するパケットの送信経路を示す図である。

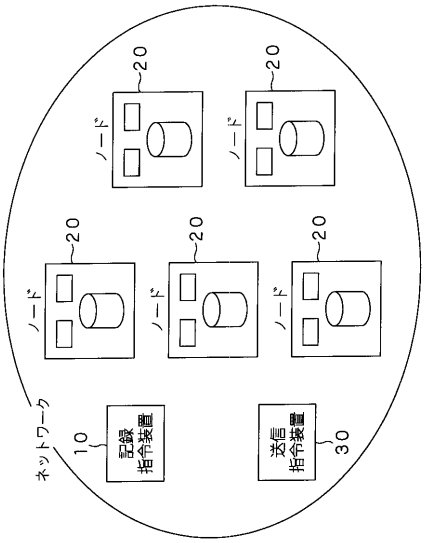
【図１１】ノードから送信指令装置に返信されるパケットの送信経路を示す図である。

【図１２】ノードから送信されたパケットを元データに復元する工程を示す図である。

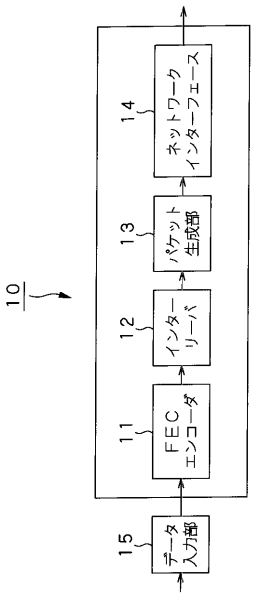
【符号の説明】

１０　記録指令装置、１１　ＦＥＣエンコーダ、１２　インターリーバ、１３パケット生成部、２０　ノード、２２　制御部、２３　記録部、３０　送信指令装置、３２　データ要求部、３３　パケット結合部、３４　デインターリーバ、３５　ＦＥＣデコーダ、４０　パケット、４２　記録確率記述部、５０　パケット、５２　返信確率記述部、５３　リクエスト記述部、６０　パケット、６２
ペイロード

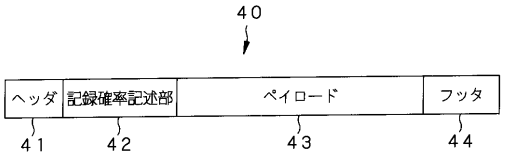
【図１】



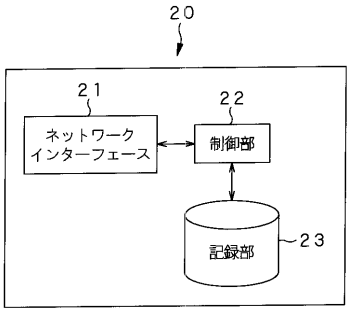
【図２】



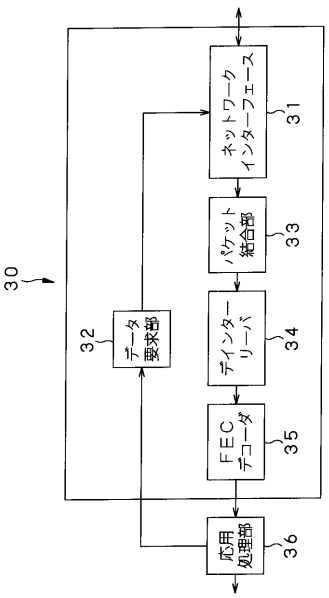
【図３】



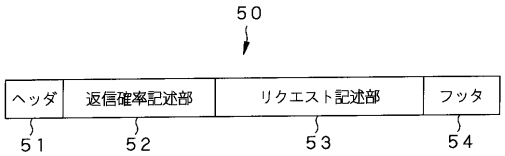
【図４】



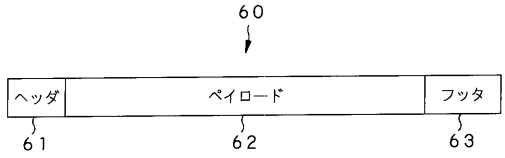
【図５】



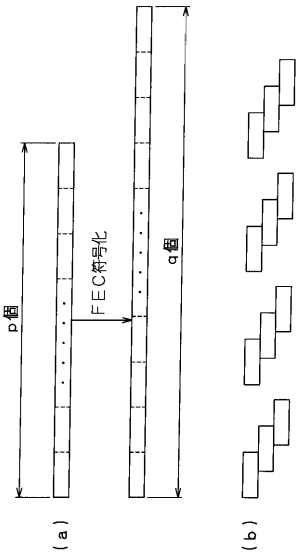
【図６】



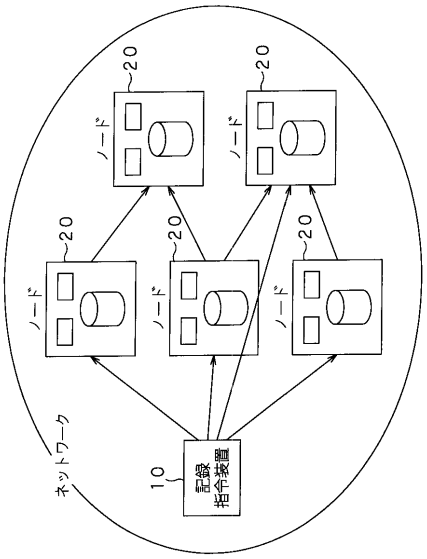
【図７】



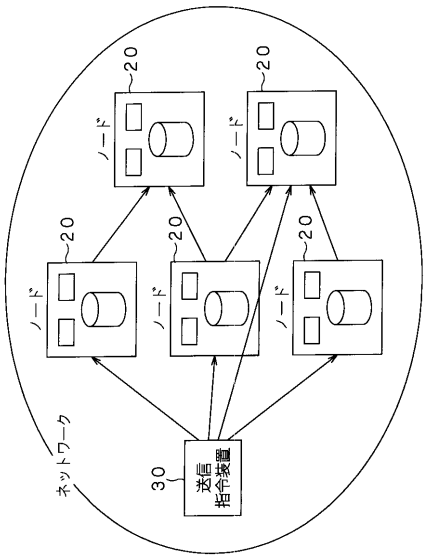
【図８】



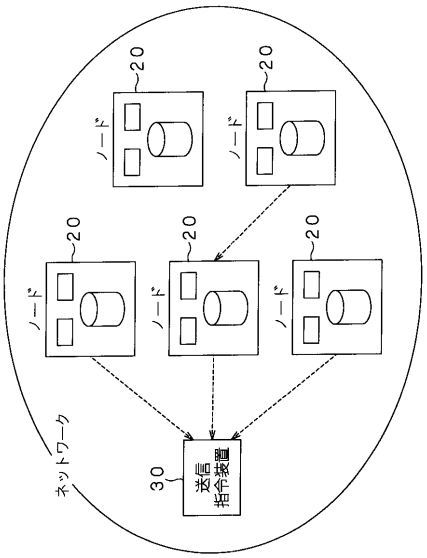
【図９】



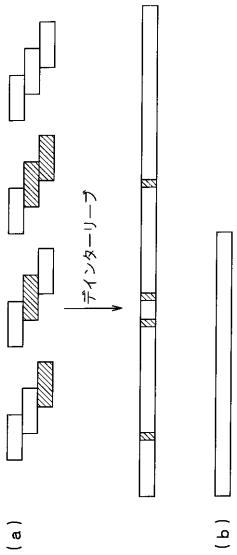
【図１０】



【図１１】



【図１２】



フロントページの続き

(72)発明者　片山　靖
東京都品川区北品川６丁目７番３５号　ソニー株式会社内

審査官　須藤　竜也

(56)参考文献　特開２０００－１５５７１２（ＪＰ，Ａ）
国際公開第９９／０３４２９１（WO，Ａ１）
国際公開第００／０３１９４５（WO，Ａ１）
特開平０３－０６４１４１（ＪＰ，Ａ）
特開平０７－８７０９２（ＪＰ，Ａ）

(58)調査した分野(Int.Cl.，ＤＢ名)
G06F　12/00
G06F　13/00
H04L　12/56